jc970 U.S. PTO

11/28/00

11-30-00

09701675 .070900

526 Rec'd PCT/PTO 28 NOV 2000

FORM PTO-1390 U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE **TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371**		ATTORNEY'S DOCKET NUMBER **PF-0531 USN** U.S. APPLICATION NO. (If known, see 37 CFR 1.5) **TO BE ASSIGNED** 09/701675
INTERNATIONAL APPLICATION NO. **PCT/US99/12906**	INTERNATIONAL FILING DATE **08 June 1999**	PRIORITY DATE CLAIMED **08 June 1998**

TITLE OF INVENTION
CELL CYCLE REGULATION PROTEINS

APPLICANT(S) FOR DO/EO/US
INCYTE PHARMACEUTICALS, INC.; BANDMAN, Olga; LAL, Preeti; TANG, Y. Tom; CORLEY, Neil C.; GUEGLER, Karl J.; BAUGHN, Mariah R.; PATTERSON, Chandra

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is the **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☐ This is an express request to promptly begin national examination procedures (35 U.S.C. 371 (f)).
4. ☐ The US has been elected by the expiration of 19 months from the priority date (PCT Article 31).
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. ☐ is attached hereto (required only if not communicated by the International Bureau)
   b. ☐ has been communicated by the International Bureau.
   c. ☒ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
7. ☒ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. ☐ are attached hereto (required only if not communicated by the International Bureau).
   b. ☐ have been communicated by the International Bureau.
   c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
   d. ☒ have not been made and will not be made.
8. ☐ An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☐ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ An English language translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 16 below concern document(s) or information included:

11. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.27 and 3.31 is included.
13. ☐ A FIRST preliminary amendment.
    ☐ A SECOND or SUBSEQUENT preliminary amendment.
14. ☐ A substitute specification.
15. ☐ A change of power of attorney and/or address letter.
16. ☒ Other items or information:

1) Transmittal Letter (2 pp, in duplicate)
2) Return Postcard
3) Express Mail Label No.: <u>EL 579 976 080 US</u>

09701675 .070901

525 Rec'd PCT/PTO 28 NOV 2000

U.S. APPLICATION NO. (if known, see 37 CFR 1.5) TO BE ASSIGNED 09/701675	INTERNATIONAL APPLICATION NO.: PCT/US99/12906	ATTORNEY'S DOCKET NUMBER PF-0531 USN

17. ☐ The following fees are submitted:

BASIC NATIONAL FEE (37 CFR 1.492(a)(1)-(5):

Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO.......**$1000.00**

International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO..**$860.00**

International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO........**$710.00**

☒International preliminary examination fee paid to USPTO (37 CFR 1.482) but all claims did not satisfy provisions of PCT Article 33(1)-(4)..........**$690.00**

International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4)..................**$100.00**

ENTER APPROPRIATE BASIC FEE AMOUNT =				$690.00	
Surcharge of **$130.00** for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).				$	
CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total Claims	23 =	3	X $ 18.00	$54.00	
Independent Claims	2 =		X $ 80.00	$	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ $270.00	$	
TOTAL OF ABOVE CALCULATIONS =				$744.00	
☐ Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2.				$	
SUBTOTAL =				$744.00	
Processing fee of **$130.00** for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest clailmed priority date (37 CFR 1492(f)). +				$	
TOTAL NATIONAL FEE =				$744.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by the appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00** per property +				$	
TOTAL FEES ENCLOSED =				$744.00	
				Amount to be Refunded:	$
				Charged:	$

a. ☐ A check in the amount of $________ to cover the above fees is enclosed.
b. ☒ Please charge my Deposit Account No. 09-0108 in the amount of $ 744.00 to cover the above fees.
c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 09-0108. A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

INCYTE GENOMICS, INC.
3160 Porter Drive
Palo Alto, CA 94304

[signature]
SIGNATURE

NAME: Diana Hamlet-Cox

REGISTRATION NUMBER: 33,302

DATE: 28 November 2000

CELL CYCLE REGULATION PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of cell cycle regulation proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative and immune disorders.

BACKGROUND OF THE INVENTION

Cell division is the fundamental process by which all living things grow and reproduce. In unicellular organisms such as yeast and bacteria, each cell division doubles the number of organisms, while in multicellular species many rounds of cell division are required to replace cells lost by wear or by programmed cell death, and for cell differentiation to produce a new tissue or organ. Details of the cell division cycle may vary, but the basic process consists of three principle events. The first event, interphase, involves preparations for cell division, replication of the DNA, and production of essential proteins. In the second event, mitosis, the nuclear material is divided and separates to opposite sides of the cell. The final event, cytokinesis, is division and fission of the cell cytoplasm. The sequence and timing of cell cycle transitions is under the control of the cell cycle regulation system which controls the process by positive or negative regulatory circuits at various check points.

Regulated progression of the cell cycle depends on the integration of growth control pathways with the basic cell cycle machinery. Cell cycle regulators have been identified by selecting for human and yeast cDNAs that block or activate cell cycle arrest signals in the yeast mating pheromone pathway when they are over expressed. Known regulators include human CPR (cell cycle progression restoration) genes, such as CPR8 and CPR2, and yeast CDC (cell division control) genes, including CDC91, that block the arrest signals. The CPR genes express a variety of proteins including cyclins, tumor suppressor binding proteins, chaperones, transcription factors, translation factors, and RNA-binding proteins (Edwards, M. C. et al.(1997) Genetics 147:1063-1076.)

Several cell cycle transitions, including the entry and exit of a cell from mitosis, are dependent upon the activation and inhibition of cyclin-dependent kinases (Cdks). The

Cdks are composed of a kinase subunit, Cdk, and an activating subunit, cyclin, in a complex that is subject to many levels of regulation. There appears to be a single Cdk in S. cerevisiae and S. pombe whereas mammals have a variety of specialized Cdks. Cyclins act by binding to and activating cyclin-dependent protein kinases which then phosphorylate and activate selected proteins involved in the mitotic process. The Cdk-cyclin complex is both positively and negatively regulated by phosphorylation, and by targeted degradation involving molecules such as CDC4 and CDC53. In addition, Cdks are further regulated by binding to inhibitors and other proteins such as Suc1 that modify their specificity or accessibility to regulators. (Patra, D., and Dunphy, W. G. (1996) Genes Dev. 10:1503-1515; and Mathias, N. et al. (1996) Mol. Cell Biol. 16: 6634-6643.)

Modulation of mRNA processing plays an important role during the cell cycle as many transcripts are differentially regulated in distinct cell types and tissues. mRNA processing includes differential, or alternate splicing of exons, which allows the expression of functionally distinct proteins from similar mRNAs. Both constitutive and alternative splicing occurs on spliceosomes, which are complex particles composed of small nuclear ribonucleoproteins (snRNPs) and non-snRNP proteins. The SR (serine- and arginine-rich) family of non-snRNP splicing factors are required at early stages of spliceosome assembly, have distinct but overlapping specificities for different pre-mRNAs, and can alter splice site choice. Several of the mammalian SR proteins including HRS, ASF/SF2, and SC-35, have been characterized. The sequence of HRS is related to the human ASF/SF2 protein and the Drosophila B52/SRp55 protein. The latter has been shown to function in vitro to regulate 5-prime splice site selection. (Diamond, R. H. et al. (1993) J. Biol. Chem. 268:15185-15192.)

Cell cycle progression also requires coordinated expression of other proteins. Blimp-1 regulatory factor appears during the differentiation of B lymphocytes into immunoglobulin secretory cells. Blimp-1 contains Kruppel-type zinc finger motifs and proline-rich and acidic regions similar to those of known transcription factors. Stable or transient transfection of Blimp-1 into B cell lymphoma lines leads to the expression of many of the phenotypic changes associated with B cell differentiation into an early plasma cell stage, including induction of J chain message and immunoglobulin secretion,

up-regulation of syndecan-1, and increased cell size and granularity. Inhibin, a member of the transforming growth factor beta-like family has been shown to modulate cell differentiation and proliferation by antagonizing the DNA synthesis activity of activin. (Turner, C.A. et al. (1994) Cell (1994) 77:297-306; and Matzuk, M. M. et al. (1994) Proc. Nat. Acad. Sci. 91:8817-8821.)

Progression through the cell cycle and consequently, cell proliferation, is governed by the complex interactions of protein complexes composed of cyclins, cyclin-dependent protein kinases, cell growth factors, binding of inhibitory polypeptides, and associated proteins. Cancers and immune disorders are characterized by uncoordinated cell proliferation. Certain cell proliferation disorders can be identified by changes in the protein complexes that normally control progression through the cell cycle. A primary treatment strategy involves reestablishing control over cell cycle progression by manipulation of the proteins involved in cell cycle regulation. (Nigg, E.A. (1995) BioEssays 17:471-480.)

The discovery of new cell cycle regulation proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, treatment, and prevention of cell proliferative and immune disorders.

SUMMARY OF THE INVENTION

The invention features substantially purified polypeptides, cell cycle regulation proteins, referred to collectively as "CECRP" and individually as "CECRP-1", "CECRP-2", "CECRP-3", "CECRP-4", and "CECRP-5". In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, (SEQ ID NO:1 through 5) and fragments thereof.

The invention further provides a substantially purified variant having at least 90% amino acid identity to the amino acid sequences of SEQ ID NO:1 through 5, and fragments thereof. The invention also provides an isolated and purified polynucleotide

encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1 through 5, and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1 through 5, and fragments thereof.

Additionally, the invention provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting SEQ ID NO:1 through 5, and fragments thereof, as well as an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1 through 5, and fragments thereof.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, (SEQ ID NO:6 through 10), and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6 through 10, and fragments thereof as well as an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6 through 10, and fragments thereof.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1 through 5, and fragments thereof. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1 through 5, and fragments thereof, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide

encoding the polypeptide under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1 through 5, and fragments thereof in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1 through 5, and fragments thereof, as well as a purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a cell proliferative disorder, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1 through 5, and fragments thereof in conjunction with a suitable pharmaceutical carrier.

The invention also provides a method for treating or preventing a cell proliferative disorder, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1 through 5 and fragments thereof.

The invention also provides a method for treating or preventing an immune disorder, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1 through 5, and fragments thereof in conjunction with a suitable pharmaceutical carrier.

The invention also provides a method for treating or preventing an immune disorder, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1 through 5, and fragments thereof.

The invention also provides a method for detecting a polynucleotide encoding the

polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1 through 5, and fragments thereof in a biological sample containing nucleic acids, the method comprising the steps of: (a) hybridizing the complement of the polynucleotide sequence encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1 through 5, and fragments thereof to at least one of the nucleic acids of the biological sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide encoding the polypeptide in the biological sample. In one aspect, the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to the hybridizing step.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications and which might be used in

connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"CECRP," as used herein, refers to the amino acid sequences of substantially purified CECRP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist," as used herein, refers to a molecule which, when bound to CECRP, increases or prolongs the duration of the effect of CECRP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of CECRP.

An "allelic variant," as this term is used herein, is an alternative form of the gene encoding CECRP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding CECRP, as described herein, include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same as CECRP or a polypeptide with at least one functional characteristic of CECRP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CECRP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CECRP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent CECRP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity,

hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CECRP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" or "amino acid sequence," as used herein, refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments," "immunogenic fragments," or "antigenic fragments" refer to fragments of CECRP which are preferably about 5 to about 15 amino acids in length, most preferably 14 amino acids, and which retain some biological activity or immunological activity of CECRP. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification," as used herein, relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art. (See, e.g., Dieffenbach, C.W. and G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, pp.1-5.)

The term "antagonist," as it is used herein, refers to a molecule which, when bound to CECRP, decreases the amount or the duration of the effect of the biological or immunological activity of CECRP. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of CECRP.

As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind CECRP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized

chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant," as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense," as used herein, refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

As used herein, the term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic CECRP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity," as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular

importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition comprising a given amino acid sequence," as these terms are used herein, refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation, an aqueous solution, or a sterile composition. Compositions comprising polynucleotide sequences encoding CECRP or fragments of CECRP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts, e.g., NaCl, detergents, e.g.,sodium dodecyl sulfate (SDS), and other components, e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.

"Consensus sequence," as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using XL-PCR™ (Perkin Elmer, Norwalk, CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW™ Fragment Assembly system (GCG, Madison, WI). Some sequences have been both extended and assembled to produce the consensus sequence.

As used herein, the term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding CECRP, by Northern analysis is indicative of the presence of nucleic acids encoding CECRP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding CECRP.

A "deletion," as the term is used herein, refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative," as used herein, refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at

least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

The term "similarity," as used herein, refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MegAlign™ program (DNASTAR, Inc., Madison WI). The MegAlign™ program can create alignments between two or more sequences according to different methods, e.g., the clustal method. (See, e.g., Higgins, D.G. and P.M. Sharp (1988) Gene 73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and

sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) Methods Enzymol. 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs), as described herein, are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355.)

The term "humanized antibody," as used herein, refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization," as the term is used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

As used herein, the term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" or "addition," as used herein, refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray," as used herein, refers to an arrangement of distinct

polynucleotides arrayed on a substrate, e.g., paper, nylon or any other type of membrane, filter, chip, glass slide, or any other suitable solid support.

The terms "element" or "array element" as used herein in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate," as it appears herein, refers to a change in the activity of CECRP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CECRP.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which, comprise a region of unique polynucleotide sequence that specifically identifies a polynucleotide of SEQ ID NO:6-10, for example, as distinct from any other sequence in the same genome. For example, a fragment of a polynucleotide of SEQ ID NO:6-10 is useful in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:6-10 from related polynucleotide sequences. A fragment of a polynucleotide of SEQ ID NO:6-10 is at least about 15-20 nucleotides in length. The precise length of the fragment of polynucleotide of SEQ ID NO:6-10 and the region of a polynucleotide of SEQ ID NO:6-10 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment. Alternatively, a fragment when translated, would produce polypeptides retaining some functional characteristic, e.g., antigenicity, or structural domain characteristic, e.g., ATP-binding site, of the full-length polypeptide.

The terms "operably associated" or "operably linked," as used herein, refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the translation of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the sequence encoding the polypeptide but still bind to

operator sequences that control expression of the polypeptide.

The term “oligonucleotide,” as used herein, refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. As used herein, the term “oligonucleotide” is substantially equivalent to the terms "amplimer," "primer," "oligomer," and "probe," as these terms are commonly defined in the art.

“Peptide nucleic acid” (PNA), as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell. (See, e.g., Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8:53-63.)

The term “sample,” as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acids encoding CECRP, or fragments thereof, or CECRP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a solid support; a tissue; a tissue print; etc.

As used herein, the terms “specific binding” or “specifically binding” refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope “A,” the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

As used herein, the term “stringent conditions” refers to conditions which permit hybridization between polynucleotides and the claimed polynucleotides. Stringent conditions can be defined by salt concentration, the concentration of organic solvent (e.g., formamide), temperature, and other conditions well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the

concentration of formamide, or raising the hybridization temperature.

The term "substantially purified," as used herein, refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution," as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation," as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of CECRP, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE™ software.

The term "variant," when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to CECRP. This definition may also include, for example, "allelic" (as defined above), "splice," "species," or "polymorphic" variants. A splice variant may have significant identity to a reference molecule, but will

generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

THE INVENTION

The invention is based on the discovery of new cell cycle regulation proteins (CECRP), the polynucleotides encoding CECRP, and the use of these compositions for the diagnosis, treatment, or prevention of diseases associated with cell proliferation and immune disorders.

In Table 1, columns 1 and 2 show the sequence identifiers for each of the amino acid and nucleic acid sequence, respectively. Column 3 shows the Clone ID of the Incyte Clone in which nucleic acids encoding each CECRP were first identified, and column 4, the cDNA library of this clone. Column 5 is entitled fragments, and shows the Incyte clones (and libraries) and shotgun sequences useful as fragments, for example, in hybridization technologies, and which are part of the consensus nucleotide sequence of each CECRP. The columns of Table 2 show various properties of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues; column 3, potential phosphorylation sites; column 4, potential glycosylation sites; column 5, identifying sequences and/or structural motifs; and column 6, analytical methods used to identify the protein through sequence homologies and protein motifs.

The columns of Table 3 show the tissue expression of each nucleic acid sequence by Northern analysis, diseases or disorders associated with this tissue expression, and the vector into which each cDNA was cloned.

The invention also encompasses CECRP variants. A preferred CECRP variant is

one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the CECRP amino acid sequence, and which contains at least one functional or structural characteristic of CECRP.

The invention also encompasses polynucleotides which encode CECRP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising the sequence selected from the group consisting of SEQ ID NO:6 through 10.

The invention also encompasses a variant of a polynucleotide sequence encoding CECRP. In particular, such a variant polynucleotide sequence will have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding CECRP. A particular aspect of the invention encompasses a variant of a nucleic acid sequence selected from the group consisting of SEQ ID NO:6 through 10 which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:6 through SEQ ID NO:10. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CECRP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CECRP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring CECRP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode CECRP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring CECRP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CECRP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a

particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CECRP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode CECRP and CECRP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CECRP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:6 through 10, or a fragments thereof, under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.)

Methods for DNA sequencing and analysis are well known and generally available in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE® (Amersham Pharmacia Biotech, Piscataway NJ), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE Amplification System (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the ABI Catalyst 800 (Perkin Elmer) or a Hamilton Micro Lab 2200 (Hamilton, Reno NV) in combination with thermal cyclers (for PCR). Sequencing is then carried out using either ABI 373 or 377 DNA sequencers (Perkin Elmer) or the MEGABASE capillary electrophoresis (Molecular Dynamics), and the sequences are analyzed using tools (computer programs and algorithms) which are well known in the art. (Ausubel (1997, unit 7.7); Meyers, R.A. (1995; Molecular Biology and Biotechnology, Wiley VCH, Inc, New York NY, p 856-853).

The nucleic acid sequences encoding CECRP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-306). Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO™ 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic

separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., Genotyper™ and Sequence Navigator™, Perkin Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode CECRP may be cloned in recombinant DNA molecules that direct expression of CECRP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express CECRP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter CECRP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding CECRP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232.) Alternatively, CECRP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A Peptide Synthesizer (Perkin Elmer). Additionally, the amino acid sequence of CECRP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g, Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman and Co., New York, NY.)

In order to express a biologically active CECRP, the nucleotide sequences encoding CECRP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding CECRP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CECRP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CECRP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding CECRP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, ch. 4, 8, and 16-17; and Ausubel, F.M. et al. (1995, and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding CECRP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV)) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding CECRP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CECRP can be achieved using a multifunctional E. coli vector such as Bluescript® (Stratagene) or pSport1™ plasmid (GIBCO BRL). Ligation of sequences encoding CECRP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of CECRP are needed, e.g. for the production of antibodies, vectors which direct high level expression of CECRP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CECRP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, supra; and Grant et al. (1987) Methods Enzymol. 153:516-54; Scorer, C. A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of CECRP. Transcription of sequences encoding CECRP may be driven viral promoters, e.g., the 35S and 19S

promoters of CaMV used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) EMBO J. 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., Hobbs, S. or Murry, L.E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CECRP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CECRP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

For long term production of recombinant proteins in mammalian systems, stable expression of CECRP in cell lines is preferred. For example, sequences encoding CECRP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ or *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; and Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14; and Murry, supra.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP) (Clontech, Palo Alto, CA), ß glucuronidase and its substrate ß-D-glucuronoside, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CECRP is inserted within a marker gene sequence, transformed cells containing sequences encoding CECRP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CECRP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding CECRP and that express CECRP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of CECRP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CECRP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Section IV; Coligan, J. E. et al. (1997 and periodic supplements) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York, NY; and Maddox, D.E. et al. (1983) J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CECRP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding CECRP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Pharmacia & Upjohn (Kalamazoo, MI), Promega (Madison, WI), and U.S. Biochemical Corp. (Cleveland, OH). Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding CECRP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CECRP may be designed to contain signal sequences which direct secretion of CECRP

through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Bethesda, MD) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CECRP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CECRP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CECRP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CECRP encoding sequence and the heterologous protein sequence, so that CECRP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel, F. M. et al. (1995 and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch 10. A variety of commercially available

kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled CECRP may be achieved in vitro using the TNT™ rabbit reticulocyte lysate or wheat germ extract systems (Promega, Madison, WI). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, preferably ^{35}S-methionine.

Fragments of CECRP may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of CECRP may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural homology exists between CECRP and cell cycle regulation proteins. In addition the expression of CECRP is closely associated with cell proliferation. Therefore, in cancers or immune disorders where CECRP is an activator, or enhancer, and is promoting cell proliferation, it is desirable to decrease the expression of CECRP. In cancers or immune disorders where CECRP is an inhibitor or suppressor of cellular processes that modulate cell proliferation it is desirable to provide the protein or to increase the expression of CECRP.

In one embodiment, where CECRP is an inhibitor of cell proliferation, CECRP or a fragment or derivative thereof may be administered to a subject to treat or prevent a cell proliferation disorder. Such a disorder may include, but is not limited to, actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas,

parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a pharmaceutical composition comprising purified CECRP may be used to treat or prevent a cell proliferative disorder including, but not limited to, those listed above.

In another embodiment, an agonist which is specific for CECRP may be administered to a subject to treat or prevent a cell proliferative disorder including, but not limited to, those cancers listed above.

In another further embodiment, a vector capable of expressing CECRP, or a fragment or a derivative thereof, may be administered to a subject to treat or prevent a cell proliferative disorder including, but not limited to, those listed above.

In a further embodiment where CECRP is an inhibitor of cell proliferation, CECRP or a fragment or derivative thereof may be administered to a subject to treat or prevent an immune disorder. Such a disorder may include, but is not limited to, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma.

In another embodiment, a pharmaceutical composition comprising purified CECRP may be used to treat or prevent an immune disorder including, but not limited to, those listed above.

In another embodiment, an agonist which is specific for CECRP may be administered to a subject to treat or prevent an immune disorder including, but not limited to, those listed above.

In another further embodiment, a vector capable of expressing CECRP, or a fragment or a derivative thereof, may be administered to a subject to treat or prevent an immune disorder including, but not limited to, those listed above.

In yet another embodiment, where CECRP is a promoter of cell proliferation, antagonists which decrease the expression or activity of CECRP may be administered to a subject to treat or prevent a cell proliferation disorder including, but not limited to, those listed above. In one aspect, antibodies which specifically bind CECRP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express CECRP.

In another embodiment, a vector expressing the complement of the polynucleotide encoding CECRP may be administered to a subject to treat or prevent a cell proliferation disorder including, but not limited to, those listed above.

In yet another embodiment, where CECRP is a promoter of cell proliferation, antagonists which decrease the expression or activity of CECRP may be administered to a subject to treat or prevent an immune disorder including, but not limited to, those listed above.

In another embodiment, a vector expressing the complement of the polynucleotide encoding CECRP may be administered to a subject to treat or prevent an immune disorder including, but not limited to, those listed above.

In one further embodiment, CECRP or a fragment or derivative thereof may be added to cells to stimulate cell proliferation. In particular, CECRP may be added to a cell in culture or cells in vivo using delivery mechanisms such as liposomes, viral based vectors, or electroinjection for the purpose of promoting cell proliferation and tissue or organ regeneration. Specifically, CECRP may be added to a cell, cell line, tissue or organ culture in vitro or ex vivo to stimulate cell proliferation for use in heterologous or autologous transplantation. In some cases, the cell will have been preselected for its ability to fight an infection or a cancer or to correct a genetic defect in a disease such as sickle cell anemia, β thalassemia, cystic fibrosis, or Huntington's chorea.

In another embodiment, an agonist which is specific for CECRP may be administered to a cell to stimulate cell proliferation, as described above.

In another embodiment, a vector capable of expressing CECRP, or a fragment or a derivative thereof, may be administered to a cell to stimulate cell proliferation, as described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CECRP may be produced using methods which are generally known in the art. In particular, purified CECRP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CECRP. Antibodies to CECRP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with CECRP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to CECRP have an amino acid sequence consisting of at least about 5 amino

acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of CECRP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CECRP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CECRP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; and Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for CECRP may also be generated. For example, such fragments include, but are not limited to, F(ab')2 fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression

libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CECRP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CECRP epitopes is preferred, but a competitive binding assay may also be employed. (Maddox, supra.)

In another embodiment of the invention, the polynucleotides encoding CECRP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding CECRP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding CECRP. Thus, complementary molecules or fragments may be used to modulate CECRP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CECRP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding CECRP. (See, e.g., Sambrook, supra; and Ausubel, supra.)

Genes encoding CECRP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding CECRP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous

nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding CECRP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding CECRP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be

prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding CECRP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of CECRP, antibodies to CECRP, and mimetics, agonists, antagonists, or inhibitors of CECRP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be

administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer

solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic

solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of CECRP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example CECRP or fragments thereof, antibodies of CECRP, and agonists, antagonists or inhibitors of CECRP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of therapeutic to toxic effects is the therapeutic index, and it can be expressed as the ED_{50}/LD_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related

to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind CECRP may be used for the diagnosis of disorders characterized by expression of CECRP, or in assays to monitor patients being treated with CECRP or agonists, antagonists, or inhibitors of CECRP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for CECRP include methods which utilize the antibody and a label to detect CECRP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring CECRP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of CECRP expression. Normal or standard values for CECRP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to CECRP under conditions suitable for complex formation The

amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of CECRP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding CECRP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of CECRP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of CECRP, and to monitor regulation of CECRP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CECRP or closely related molecules may be used to identify nucleic acid sequences which encode CECRP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding CECRP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the CECRP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequences of SEQ ID NO:6 through SEQ ID NO:10 or from genomic sequences including promoters, enhancers, and introns of the CECRP gene.

Means for producing specific hybridization probes for DNAs encoding CECRP include the cloning of polynucleotide sequences encoding CECRP or CECRP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase

coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CECRP may be used for the diagnosis of a disorder associated with expression of CECRP. Examples of such a disorder include, but are not limited to, actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma. The polynucleotide sequences encoding CECRP may be used in Southern or Northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and ELISA assays; and in microarrays utilizing fluids or tissues from patients to detect altered CECRP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding CECRP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding CECRP may be labeled by standard methods and

added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding CECRP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CECRP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CECRP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding CECRP may involve the use of PCR. These oligomers may be chemically

synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding CECRP, or a fragment of a polynucleotide complementary to the polynucleotide encoding CECRP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of CECRP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; and Duplaa, C. et al. (1993) Anal. Biochem. 229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding CECRP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial

chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, R.A. (ed.) Molecular Biology and Biotechnology, VCH Publishers New York, NY, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding CECRP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CECRP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CECRP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with CECRP, or fragments thereof, and washed. Bound CECRP is then detected by methods well known in the art. Purified CECRP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding CECRP specifically compete with a test compound for binding CECRP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CECRP.

In additional embodiments, the nucleotide sequences which encode CECRP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications, mentioned above and below, in particular U.S. Ser. No. 60/088,695, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech (Palo Alto, CA) or isolated at Incyte from tissues described in Table 4. The tissue was homogenized and lysed in guanidinium isothiocyanate, and the lysate was centrifuged over a CsCl cushion. Alternatively, the

tissue was homogenized and lysed in phenol or a suitable mixture of denaturants such as TRIzol reagent (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate, and the lysate was extracted with chloroform (1:5 v/v). RNA was precipitated from lysates with either isopropanol or sodium acetate and ethanol. Alternatively, RNA was purified from lysates by preparative agarose gel electrophoresis and recovered from Whatman P81 paper (Whatman, Lexington, MA). Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity, and RNA was maintained in RNase-free solutions. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega, Madison, WI), Oligotex resin, or the Oligotex kit (QIAGEN Inc, Chatsworth, CA). Alternatively, RNA was isolated directly from tissue lysates using the RNA Isolation kit (Stratagene) or the Ambion PolyA Quick kit (Ambion, Austin, TX).

RNA was used for cDNA synthesis and construction of the cDNA libraries according to procedures recommended in the UNIZAP vector (Stratagene, La Jolla, CA) or SuperScript plasmid system (Life Technologies, Inc), both of which are based on methods well known in the art (Ausubel, 1997, units 5.1-6.6). Alternatively, cDNA libraries were constructed by Stratagene using RNA provided by Incyte. Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and cDNA was digested with an appropriate restriction enzyme(s). For most libraries, cDNA was size-selected (300-1000 bp) using Sephacryl S1000 or Sepharose CL2B or CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., pBluescript (Stratagene), pSPORT 1 (Life Technologies), pINCY (Incyte Pharmaceuticals Inc, Palo Alto, CA). pINCY was amplified in JM109 cells and purified using the QiaQuick column (QIAGEN Inc). Recombinant plasmids were transformed into competent E. coli cells, e.g., XL1-Blue, XL1-BlueMRF, or SOLR (Stratagene) or DH5α, DH10B, or ElectroMAX DH10B (Life Technologies).

II. Isolation of cDNA Clones

Plasmids were recovered from host cells by in vivo excision (UniZAP vector

system, Stratagene) or by cell lysis. Plasmids were purified using the MAGIC MINIPREPS DNA purification system (Promega, Madison, WI); Miniprep kit (Advanced Genetic Technologies Corporation, Gaithersburg, MD); QIAwell-8 Plasmid, QIAwell PLUS DNA, or QIAwell ULTRA DNA purification systems; or REAL Prep 96 plasmid kit (QIAGEN Inc) using the recommended protocol. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR (Rao, V.B. (1994) Anal. Biochem. 216:1-14) in a high-throughput format. Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates (Genetix Ltd, Christchurch UK) and concentration of amplified plasmid DNA was quantified fluorometrically using Pico Green Dye (Molecular Probes, Eugene OR) and a Fluoroscan II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

The cDNAs were prepared for sequencing using either an ABI Catalyst 800 (Perkin Elmer) or a Hamilton Micro Lab 2200 (Hamilton, Reno, NV) in combination with Peltier Thermal Cyclers (PTC200; MJ Research, Watertown MA). The cDNAs were sequenced on the ABI 373 or 377 DNA Sequencing systems (Perkin Elmer) by the method of Sanger F and A.R. Coulson (1975; J. Mol. Biol. 94:441-448) using standard ABI protocols, base calling software, and kits. Alternatively, cDNAs were sequenced using solutions and dyes from Amersham Pharmacia Biotech. Reading frame was determined using standard methods (Ausubel, supra).

The cDNA sequences presented in Table 1 and the full length nucleotide and amino acid sequences disclosed in the Sequence Listing were queried against databases such as GenBank primate (pri), rodent (rod), mammalian (mamp), vertebrate (vrtp), and eukaryote (eukp) databases, SwissProt, BLOCKS, and other databases which contain previously identified and annotated motifs and sequences. Algorithms such as Smith Waterman which deal with primary sequence patterns and secondary structure gap penalties (Smith, T. et al. (1992) Protein Engineering 5:35-51) and programs and

algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S.F. (1993) J. Mol. Evol 36:290-300; and Altschul et al. (1990) J. Mol. Biol. 215:403-410), and HMM (Hidden Markov Models; Eddy, S.R. (1996) Cur. Opin. Str. Biol. 6:361-365 and Sonnhammer, E.L.L. et al. (1997) Proteins 28:405-420) were used to assemble and analyze nucleotide and amino acid sequences. The databases, programs, algorithms, methods and tools are available, well known in the art, and described in Ausubel (supra, unit 7.7), in Meyers, R.A. (1995; Molecular Biology and Biotechnology, Wiley VCH, Inc, New York NY, p 856-853), in documentation provided with software (Genetics Computer Group (GCG), Madison WI), and on the world wide web (www). Two comprehensive websites which list, describe, and/or link many of the databases and tools are: 1) the www resource in practical sequence analysis (http://genome.wustl.edu/eddy/bio5495/online_resources.html), and 2) the bibliography of computational gene recognition (http://linkage.rockefeller.edu/wli/gene/ programs.html). For example, the first website

links PFAM as a database (http://genome.wustl. edu/Pfam/) and as an HMM search tool (http://genome.wustl.edu/eddy/cgi-bin/hmm_page.cgi).

Table 5 summarizes the databases and tools used herein. The first column of Table 5 shows the tool, program, or algorithm; the second column, the database; the third column, a brief description; and the fourth column (where applicable), scores for determining the strength of a match between two sequences (the higher the value, the more homologous).

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; and Ausubel, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST are used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ™ database (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based

hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar.

The basis of the search is the product score, which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of Northern analysis are reported as a list of libraries in which the transcript encoding CECRP occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

V. Extension of CECRP Encoding Polynucleotides

The nucleic acid sequences of Incyte Clones 037377, 162871, 236062, 1596581, and 1853196, were used to design oligonucleotide primers for extending partial nucleotide sequences to full length. For each nucleic acid sequence, one primer was synthesized to initiate extension of an antisense polynucleotide, and the other was synthesized to initiate extension of a sense polynucleotide. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO™ 4.06 (National Biosciences, Plymouth, MN), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (GIBCO BRL) were used to extend the sequence. If more than one extension is necessary or desired, additional sets of primers are designed

to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR™ kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix. PCR was performed using the Peltier Thermal Cycler (PTC200; M.J. Research, Watertown, MA), beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, with the following parameters:

Step 1	94° C for 1 min (initial denaturation)
Step 2	65° C for 1 min
Step 3	68° C for 6 min
Step 4	94° C for 15 sec
Step 5	65° C for 1 min
Step 6	68° C for 7 min
Step 7	Repeat steps 4 through 6 for an additional 15 cycles
Step 8	94° C for 15 sec
Step 9	65° C for 1 min
Step 10	68° C for 7:15 min
Step 11	Repeat steps 8 through 10 for an additional 12 cycles
Step 12	72° C for 8 min
Step 13	4° C (and holding)

A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6% to 0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQUICK™ (QIAGEN Inc.), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13 μl of ligation buffer, 1μl T4-DNA ligase (15 units) and 1μl T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2 to 3 hours, or overnight at 16° C. Competent E. coli cells (in 40 μl of appropriate media) were transformed with 3 μl of ligation mixture and cultured in 80 μl of SOC medium. (See, e.g., Sambrook, supra, Appendix A, p. 2.) After incubation for one hour at 37°C, the E. coli mixture was plated

on Luria Bertani (LB) agar (See, e.g., Sambrook, supra, Appendix A, p. 1) containing carbenicillin (2x carb). The following day, several colonies were randomly picked from each plate and cultured in 150 μl of liquid LB/2x carb medium placed in an individual well of an appropriate commercially-available sterile 96-well microtiter plate. The following day, 5 μl of each overnight culture was transferred into a non-sterile 96-well plate and, after dilution 1:10 with water, 5 μl from each sample was transferred into a PCR array.

For PCR amplification, 18 μl of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

Step 1	94° C for 60 sec
Step 2	94° C for 20 sec
Step 3	55° C for 30 sec
Step 4	72° C for 90 sec
Step 5	Repeat steps 2 through 4 for an additional 29 cycles
Step 6	72° C for 180 sec
Step 7	4° C (and holding)

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequence of SEQ ID NO:6 through 10 to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:6 through 10 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-

art software such as OLIGO™ 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham, Chicago, IL), and T4 polynucleotide kinase (DuPont NEN®, Boston, MA). The labeled oligonucleotides are substantially purified using a Sephadex™ G-25 superfine size exclusion dextran bead column (Pharmacia & Upjohn, Kalamazoo, MI). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba1, or Pvu II (DuPont NEN, Boston, MA).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak, Rochester, NY) is exposed to the blots to film for several hours, hybridization patterns are compared visually.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE™. Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present

invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; and Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

VIII. Complementary Polynucleotides

Sequences complementary to the CECRP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CECRP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO™ 4.06 software and the coding sequence of CECRP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CECRP-encoding transcript.

IX. Expression of CECRP

Expression and purification of CECRP is achieved using bacterial or virus-based expression systems. For expression of CECRP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express CECRP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CECRP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of

baculovirus is replaced with cDNA encoding CECRP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E. K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, CECRP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Pharmacia, Piscataway, NJ). Following purification, the GST moiety can be proteolytically cleaved from CECRP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak, Rochester, NY). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN Inc, Chatsworth, CA). Methods for protein expression and purification are discussed in Ausubel, F. M. et al. (1995 and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch 10, 16. Purified CECRP obtained by these methods can be used directly in the following activity assay.

X. Demonstration of CECRP Activity

CECRP activity is demonstrated by measuring the induction of terminal differentiation or cell cycle progression when CECRP is expressed at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT™ (Life Technologies, Gaithersburg, MD) and pCR™ 3.1 (Invitrogen, Carlsbad, CA), both of which contain the

cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP) (Clontech, Palo Alto, CA), CD64, or a CD64-GFP fusion protein. Flow cytometry detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell cycle progression or terminal differentiation. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; up or down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York, NY.

XI. Functional Assays

CECRP function is assessed by expressing the sequences encoding CECRP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT™ (Life Technologies, Gaithersburg, MD) and pCR™ 3.1 (Invitrogen, Carlsbad, CA, both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP) (Clontech, Palo Alto, CA), CD64, or a CD64-GFP

fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP, and to evaluate properties, for example, their apoptotic state. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York, NY.

The influence of CECRP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CECRP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success, NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CECRP and other genes of interest can be analyzed by Northern analysis or microarray techniques.

XII. Production of CECRP Specific Antibodies

CECRP substantially purified using polyacrylamide gel electrophoresis (PAGE)(see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the CECRP amino acid sequence is analyzed using LASERGENE™ software (DNASTAR Inc.) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of

appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel supra, ch. 11.)

Typically, oligopeptides 15 residues in length are synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry and coupled to KLH (Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIII. Purification of Naturally Occurring CECRP Using Specific Antibodies

Naturally occurring or recombinant CECRP is substantially purified by immunoaffinity chromatography using antibodies specific for CECRP. An immunoaffinity column is constructed by covalently coupling anti-CECRP antibody to an activated chromatographic resin, such as CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing CECRP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CECRP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CECRP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CECRP is collected.

XIV. Identification of Molecules Which Interact with CECRP

CECRP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CECRP, washed, and any wells with labeled CECRP complex are assayed. Data obtained using different concentrations of CECRP are used to calculate values for the number, affinity, and association of CECRP with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	6	037377	HUVENOB01	037377H1 (HUVENOB01), 403002R6 (TMLR3DT01), 510407R6 (MPHGNOT03), 1520634F1 (BLADTUT04), 3985439H1 (UTRSTUT05)
2	7	162871	ADENINB01	162871H1 (ADENINB01), 162871X4 (ADENINB01), 162871X92 (ADENINB01), 1554775H1 (BLADTUT04)
3	8	236062	SINTNOT02	236062CT1 (SINTNOT02), 236062H1 (SINTNOT02), 1352052H1 (LATRTUT02), 2238411F6 (PANCTUT02), 3015795H1 (MUSCNOT07)
4	9	1596581	BRAINOT14	1477338F1 (CORPNOT02), 1596581F6 (BRAINOT14), 1596581H1 (BRAINOT14), 1596581T1 (BRAINOT14), 1658067H1 (URETTUT01), 2312928H1 (NGANNOT01), 3231214H1 (COTRNOT01), 3590729H1 (293TF5T01)
5	10	1853196	LUNGFET03	108390F1 (AMLBNOT01), 1211009R1 (BRSTNOT02), 1211009T1 (BRSTNOT02), 1391767F1 (THYRNOT03), 1525569F6 (UCMCL5T01), 1706512F6 (DUODNOT02), 1722946F6 (BLADNOT06), 1853196F6 (LUNGFET03)

Table 2

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Identifying sequences and/or structural motifs	Analytical Methods
1	197	S128 S15 S71 T108 T27 S71		Blimp-1	Blast
2	225	T181 S222 S122	N54	growth response protein	Blast
3	236	S44 S60 S98 S117 S123 S180 T73		Inhibin beta chain signature	Prints
4	351	T45 S165 S179 T247 T339 T90 T339 Y35 Y51	N79	cdc91	Blast
5	757	T184 S367 S6 S22 T31 S43 S81 T99 S130 S158 S162 S163 S186 S188 S271 S277 T287 T289 S323 T343 S370 T432 S488 S499 S207 S277 T298 S370 T419 S474 S488 S499 S528 T532 T539 S609 S624 Y96 Y730	N4 N63 N479 N498	cell cycle progression restoration 8 protein	Blast

Table 3

Seq ID NO:	Tissue Expression (Fraction of Total)	Disease Class (Fraction of Total)	Vector
6	Reproductive (0.294) Gastrointestinal (0.147) Hematopoietic/Immune (0.147)	Cancer (0.500) Inflammation (0.324) Fetal (0.147)	pBluescript
7	Hematopoietic/Immune (0.240) Cardiovascular (0.160) Nervous (0.160)	Cancer (0.480) Inflammation (0.240) Fetal (0.120)	pBluescript
8	Reproductive (0.302) Cardiovascular (0.186) Nervous (0.163)	Cancer (0.651) Inflammation (0.163) Other (0.140)	pBluescript
9	Gastrointestinal (0.167) Nervous (0.167) Reproductive (0.167)	Cancer (0.467) Fetal (0.233) Inflammation (0.200)	pINCY
10	Nervous (0.200) Reproductive (0.200) Cardiovascular (0.153)	Cancer (0.471) Inflammation (0.282) Other (0.106)	pINCY

Table 4

Nucleotide SEQ ID NO:	Library	Library Description
6	HUVENOB01	The HUVENOB01 library was constructed using RNA isolated from unstimulated HUV-EC-C (ATCC CRL 1730) cells. HUV-EC-C is an endothelial cell line derived from the vein of a normal human umbilical cord (ref: PNAS 81:6413).
7	ADENINB01	The ADENINB01 Library was constructed using RNA isolated from inflamed adenoid tissue of a 3-year-old child. (RNA came from Clontech.)
8	SINTNOT02	The SINTNOT02 library was constructed using RNA isolated from the small intestine of a 55-year-old Caucasian female, who died from a subarachnoid hemorrhage.
9	BRAINOT14	The BRAINOT14 Library was constructed using 1 microgram of polyA RNA isolated from brain tissue removed from the left frontal lobe of a 40-year-old female during excision of a cerebral meningeal lesion. Pathology for the associated tumor tissue indicated grade 4 gemistocytic astrocytoma.
10	LUNGFET03	The LUNGFET03 Library was constructed using 1 microgram of polyA RNA isolated from lung tissue removed from a female fetus, who died at 20 weeks' gestation.

TABLE 5

	Program/algorithm	Databases	Description	Useful Parameters
ESTs				
	Smith Waterman	GenBank	Local alignment algorithm for homology searching	min length = 49 nt <12% uncalled bases
	FASTA	GenBank	Fast nucleotide sequence database searching program for UNIX, VMS	
	BLAST	GenBank	Ultra-fast database searching program for UNIX, VMS C source	Log likelihood for exact matches is $\sim 10^{-25}$ and for homologs $>10^{-8}$
Full Length				
	Phred		Reads trace data from sequencing runs, makes base calls for assembly of cDNA sequences, produces quality scores	
	Phrap		Quality-score based assembly program for shotgun sequences	match > 56 score > 120
	CONSED		Graphical tool for editing Phrap contigs	
	GCG Assembly, Motifs, Profilescan,	GenBank PROSITE	Wisconsing PackagePrograms for the assembly, editing, and characterization of nucleotide sequences	
	Spscan		Examines proteins for secretory, signal sequences	>7 strong, 4.5-7 suggestive
	GENEMARK		Statistical analysis of nucleotide sequences to identify open reading frame	
	BLAST	GenBank SwissProt	Ultra-fast database searching program for UNIX, VMS C source	score >100, P < 1e-5
	FASTX	GenBank SwissProt	Fast amino acid sequence database searching program for UNIX, VMS	log likelihood > 17
	BLIMPS	BLOCKS PRINTS	Weighted matrix analysis for prediction of protein family	>1300 strong, 1000 - 1300 suggestive, P<1e-3
	PFAM	PROSITE	Analyses sequences 3-60 amino acids long which correspond to highly conserved regions of a protein family	Score >11 strong, 8 - 10 suggestive
	HMM		Probabilistic approaches and modeling of the primary structure of protein families	Score >11 strong, 8 - 10 suggestive
	McDNAsis Pro		Software for sequence analysis	
	LASERGENE		Software programs (EditSeq, MegAlign, PrimerSelect, Protean, SeqMan, etc.) for sequence analysis	

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, and fragments thereof.

2. A substantially purified variant having at least 90% amino acid identity to the amino acid sequence of claim 1.

3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.

4. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 3.

5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.

6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide sequence of claim 3.

7. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, and fragments thereof.

8. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 7.

9. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 7.

10. An expression vector comprising at least a fragment of the polynucleotide of claim 3.

11. A host cell comprising the expression vector of claim 10.

12. A method for producing a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, and fragments thereof, the method comprising the steps of:

a) culturing the host cell of claim 11 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

13. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.

14. A purified antibody which specifically binds to the polypeptide of claim 1.

15. A purified agonist of the polypeptide of claim 1.

16. A purified antagonist of the polypeptide of claim 1.

17. A method for treating or preventing a cell proliferation disorder, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 13.

18. A method for treating or preventing a cell proliferation disorder, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 16.

19. A method for stimulating cell proliferation, the method comprising administering to a cell an effective amount of the pharmaceutical composition of claim 13.

20. A method for treating or preventing an immune disorder, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 16.

21. A method for treating or preventing an immune disorder, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 13.

22. A method for detecting a polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, and fragments thereof in a biological sample, the method comprising the steps of:

(a) hybridizing the polynucleotide of claim 6 to at least one of the nucleic acids in the biological sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide encoding the polypeptide in the biological sample.

23. The method of claim 22 wherein the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to hybridization.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :		(11) International Publication Number:	WO 99/64593
C12N 15/12, 5/10, 1/21, C07K 14/47, 16/18, C12Q 1/68, A61K 38/17	A2	(43) International Publication Date:	16 December 1999 (16.12.99)

(21) International Application Number: PCT/US99/12906

(22) International Filing Date: 8 June 1999 (08.06.99)

(30) Priority Data:
60/088,695 8 June 1998 (08.06.98) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 60/088,695 (CIP)
Filed on 8 June 1998 (08.06.98)

(71) Applicant (*for all designated States except US*): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US).

(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: CELL CYCLE REGULATION PROTEINS

(57) Abstract

The invention provides a human cell cycle regulation proteins (CECRP) and polynucleotides which identify and encode CECRP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating or preventing disorders associated with expression of CECRP.

Docket No.: PF-0531 USN

DECLARATION AND POWER OF ATTORNEY FOR UNITED STATES PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name, and

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if more than one name is listed below) of the subject matter which is claimed and for which a United States patent is sought on the invention entitled

CELL CYCLE REGULATION PROTEINS

the specification of which:

/___/ is attached hereto.

/ X / was filed on_________ as application Serial No. _____________ and if this box contains an X /___/, was amended on _____________________________.

/ X / was filed as Patent Cooperation Treaty international application No. PCT/US99/12906 on June 8, 1999, if this box contains an X /_/, was amended on under Patent Cooperation Treaty Article 19 on ______ 2001, and if this box contains an X /_/, was amended on ______________.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge my duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim the benefit under Title 35, United States Code, §119 or §365(a)-(b) of any foreign application(s) for patent or inventor's certificate indicated below and of any Patent Cooperation Treaty international applications(s) designating at least one country other than the United States indicated below and have also identified below any foreign application(s) for patent or inventor's certificate and Patent Cooperation Treaty international application(s) designating at least one country other than the United States for the same subject matter and having a filing date before that of the application for said subject matter the priority of which is claimed:

Country	Number	Filing Date	Priority Claimed
__________	__________	__________	/_/ Yes /_/ No
__________	__________	__________	/_/ Yes /_/ No

I hereby claim the benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below.

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)
60/088,695	June 8, 1998	Expired

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner required by the first paragraph of Title 35, United States Code §112, I acknowledge my duty to disclose material information as defined in Title 37 Code of Federal Regulations, §1.56(a) which occurred between the filing date(s) of the prior application(s) and the national or Patent Cooperation Treaty international filing date of this application:

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)

I hereby appoint the following:

Lucy J. Billings	Reg. No. 36,749
Michael C. Cerrone	Reg. No. 39,132
Diana Hamlet-Cox	Reg. No. 33,302
Richard C. Ekstrom	Reg. No. 37,027
Barrie D. Greene	Reg. No. 46,740
Matthew R. Kaser	Reg. No. 44,817
Lynn E. Murry	Reg. No. 42,918
Shirley A. Recipon	Reg. No. 47,016
Susan K. Sather	Reg. No. 44,316
Michelle M. Stempien	Reg. No. 41,327
David G. Streeter	Reg. No. 43,168
Stephen Todd	Reg. No. 47,139
Christopher Turner	Reg. No. 45,167
P. Ben Wang	Reg. No. 41,420

respectively and individually, as my patent attorneys and/or agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith. Please address all communications to:

Docket No.: PF-0531 USN

LEGAL DEPARTMENT
INCYTE GENOMICS, INC.
3160 PORTER DRIVE, PALO ALTO, CA 94304

TEL: 650-855-0555 FAX: 650-849-8886 or 650-845-4166

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Sole Inventor or First Joint Inventor: 100	Full name:	OLGA BANDMAN
	Signature:	Olga Bandman
	Date:	16 February ,2001
	Citizenship:	United States
	Residence:	Mountain View, California CA
	P.O. Address:	366 Anna Avenue Mountain View, California 94043
Second Joint Inventor: 200	Full name:	PREETI LAL
	Signature:	Preeti Lal
	Date:	17th JANUARY , 2001
	Citizenship:	India
	Residence:	Santa Clara, California CA
	P.O. Address:	2382 Lass Drive Santa Clara, California 95054

Docket No.: PF-0531 USN

Third Joint Inventor:	300	Full name:	Y. TOM TANG
		Signature:	
		Date:	February 27, 2001
		Citizenship:	~~People's Republic of China~~ USA
		Residence:	CA San Jose, California
		P.O. Address:	4230 Ranwick Court San Jose, California 95118
Fourth Joint Inventor:	400	Full name:	NEIL C. CORLEY
		Signature:	
		Date:	MARCH 5, 2001
		Citizenship:	United States
		Residence:	CASTRO VALLEY, CA NCC 1/31/01 ~~Mountain View, California~~
		P.O. Address:	~~1240 Dale Avenue, #30~~ CA ~~Mountain View, California 94040~~ 20426 CROW CREEK RD CASTRO VALLEY, CA 94552 NCC 1/31/01
Fifth Joint Inventor:	500	Full name:	KARL J. GUEGLER
		Signature:	
		Date:	02/02, 2001
		Citizenship:	Switzerland
		Residence:	CA Menlo Park, California
		P.O. Address:	1048 Oakland Avenue Menlo Park, California 94025

Docket No.: PF-0531 USN

Sixth Joint Inventor:	600 Full name:	MARIAH R. BAUGHN
	Signature:	Mariah R. Bgh
	Date:	January 17, 2001
	Citizenship:	United States
	Residence:	San Leandro, California CA
	P.O. Address:	14244 Santiago Road San Leandro, California 94577

Seventh Joint Inventor:	700 Full name:	CHANDRA PATTERSON
	Signature:	Chandra Patterson
	Date:	1/17, 2001
	Citizenship:	United States CA
	Residence:	Menlo Park, California
	P.O. Address:	490 Sherwood Way, #1 Menlo Park, California 94025

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.
BANDMAN, Olga
LAL, Preeti
TANG, Y. Tom
CORLEY, Neil C.
GUEGLER, Karl J.
BAUGHN, Mariah R.
PATTERSON, Chandra

<120> CELL CYCLE REGULATION PROTEINS

<130> PF-0531 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/088,695
<151> 1998-06-08

<160> 35

<170> PERL Program

<210> 1
<211> 197
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte clone 037377

<400> 1
Met Leu Val Leu Pro Ser Pro Cys Pro Gln Pro Leu Ala Phe Ser
1 5 10 15
Ser Val Glu Thr Met Glu Gly Pro Pro Arg Arg Thr Cys Arg Ser
20 25 30
Pro Glu Pro Gly Pro Ser Ser Ser Ile Gly Ser Pro Gln Ala Ser
35 40 45
Ser Pro Pro Arg Pro Asn His Tyr Leu Leu Ile Asp Thr Gln Gly
50 55 60
Val Pro Tyr Thr Val Leu Val Asp Glu Glu Ser Gln Arg Glu Pro
65 70 75
Gly Ala Ser Gly Ala Pro Gly Gln Lys Lys Cys Tyr Ser Cys Pro
80 85 90
Val Cys Ser Arg Val Phe Glu Tyr Met Ser Tyr Leu Gln Arg His
95 100 105
Ser Ile Thr His Ser Glu Val Lys Pro Phe Glu Cys Asp Ile Cys
110 115 120
Gly Lys Ala Phe Lys Arg Ala Ser His Leu Ala Arg His His Ser
125 130 135
Ile His Leu Ala Gly Gly Gly Arg Pro His Gly Cys Pro Leu Cys
140 145 150

Pro Arg Arg Phe Arg Asp Ala Gly Glu Leu Ala Gln His Ser Arg
155 160 165
Val His Ser Gly Glu Arg Pro Phe Gln Cys Pro His Cys Pro Arg
170 175 180
Arg Phe Met Glu Gln Asn Thr Leu Gln Lys His Thr Arg Trp Lys
185 190 195
His Pro

<210> 2
<211> 225
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte clone 162871

<400> 2
Met Ala Glu Gly Glu Thr Glu Ser Pro Gly Pro Lys Lys Cys Gly
1 5 10 15
Pro Tyr Ile Ser Ser Val Thr Ser Gln Ser Val Asn Leu Met Ile
20 25 30
Arg Gly Val Val Leu Phe Phe Ile Gly Val Phe Leu Ala Leu Val
35 40 45
Leu Asn Leu Leu Gln Ile Gln Arg Asn Val Thr Leu Phe Pro Pro
50 55 60
Asp Val Ile Ala Ser Ile Phe Ser Ser Ala Trp Trp Val Pro Pro
65 70 75
Cys Cys Gly Thr Ala Ser Ala Val Ile Gly Leu Leu Tyr Pro Cys
80 85 90
Ile Asp Arg His Leu Gly Glu Pro His Lys Phe Lys Arg Glu Trp
95 100 105
Ser Ser Val Met Arg Cys Val Ala Val Phe Val Gly Ile Asn His
110 115 120
Ala Ser Ala Lys Val Asp Phe Asp Asn Asn Ile Gln Leu Ser Leu
125 130 135
Thr Leu Ala Ala Leu Ser Ile Gly Leu Trp Trp Thr Phe Asp Arg
140 145 150
Ser Arg Ser Gly Phe Gly Leu Gly Val Gly Ile Ala Phe Leu Ala
155 160 165
Thr Val Val Thr Gln Leu Leu Val Tyr Asn Gly Val Tyr Gln Tyr
170 175 180
Thr Ser Pro Asp Phe Leu Tyr Val Arg Ser Trp Leu Pro Cys Ile
185 190 195
Phe Phe Ala Gly Gly Ile Thr Met Gly Asn Ile Gly Arg Gln Leu
200 205 210
Ala Met Tyr Glu Cys Lys Val Ile Ala Glu Lys Ser His Gln Glu
215 220 225

<210> 3
<211> 236
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte clone 236062

<400> 3
Met Leu Ser Lys Gly Leu Lys Arg Lys Arg Glu Glu Glu Glu Glu
1 5 10 15
Lys Glu Pro Leu Ala Val Asp Ser Trp Trp Leu Asp Pro Gly His
20 25 30
Ala Ala Val Ala Gln Ala Pro Pro Ala Val Ala Ser Ser Ser Leu
35 40 45
Phe Asp Leu Ser Val Leu Lys Leu His His Ser Leu Gln Gln Ser
50 55 60
Glu Pro Asp Leu Arg His Leu Val Leu Val Val Asn Thr Leu Arg
65 70 75
Arg Ile Gln Ala Ser Met Ala Pro Ala Ala Ala Leu Pro Pro Val
80 85 90
Pro Ser Pro Pro Ala Ala Pro Ser Val Ala Asp Asn Leu Leu Ala
95 100 105
Ser Ser Asp Ala Ala Leu Ser Ala Ser Met Ala Ser Leu Leu Glu
110 115 120
Asp Leu Ser His Ile Glu Gly Leu Ser Gln Ala Pro Gln Pro Leu
125 130 135
Ala Asp Glu Gly Pro Pro Gly Arg Ser Ile Gly Gly Ala Ala Pro
140 145 150
Ser Leu Gly Ala Leu Asp Leu Leu Gly Pro Ala Thr Gly Cys Leu
155 160 165
Leu Asp Asp Gly Leu Glu Gly Leu Phe Glu Asp Ile Asp Thr Ser
170 175 180
Met Tyr Asp Asn Glu Leu Trp Ala Pro Ala Ser Glu Gly Leu Lys
185 190 195
Pro Gly Pro Glu Asp Gly Pro Gly Lys Glu Glu Ala Pro Glu Leu
200 205 210
Asp Glu Ala Glu Leu Asp Tyr Leu Met Asp Val Leu Val Gly Thr
215 220 225
Gln Ala Leu Glu Arg Pro Pro Gly Pro Gly Arg
230 235

<210> 4
<211> 351
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte clone 1596581

<400> 4
Met Ile Thr Asp Ala Leu Thr Ala Ile Ala Leu Tyr Phe Ala Ile
1 5 10 15
Gln Asp Phe Asn Lys Val Val Phe Lys Lys Gln Lys Leu Leu Leu

```
                20                  25                  30
Glu Leu Asp Gln Tyr Ala Pro Asp Val Ala Glu Leu Ile Arg Thr
                35                  40                  45
Pro Met Glu Met Arg Tyr Ile Pro Leu Lys Val Ala Leu Phe Tyr
                50                  55                  60
Leu Leu Asn Pro Tyr Thr Ile Leu Ser Cys Val Ala Lys Ser Thr
                65                  70                  75
Cys Ala Ile Asn Asn Thr Leu Ile Ala Phe Phe Ile Leu Thr Thr
                80                  85                  90
Ile Lys Gly Ser Ala Phe Leu Ser Ala Ile Phe Leu Ala Leu Ala
                95                 100                 105
Thr Tyr Gln Ser Leu Tyr Pro Leu Thr Leu Phe Val Pro Gly Leu
               110                 115                 120
Leu Tyr Leu Leu Gln Arg Gln Tyr Ile Pro Val Lys Met Lys Ser
               125                 130                 135
Lys Ala Phe Trp Ile Phe Ser Trp Glu Tyr Ala Met Met Tyr Val
               140                 145                 150
Gly Ser Leu Val Val Ile Ile Cys Leu Ser Phe Phe Leu Leu Ser
               155                 160                 165
Ser Trp Asp Phe Ile Pro Ala Val Tyr Gly Phe Ile Leu Ser Val
               170                 175                 180
Pro Asp Leu Thr Pro Asn Ile Gly Leu Phe Trp Tyr Phe Phe Ala
               185                 190                 195
Glu Met Phe Glu His Phe Ser Leu Phe Phe Val Cys Val Phe Gln
               200                 205                 210
Ile Asn Val Phe Phe Tyr Thr Ile Pro Leu Ala Ile Lys Leu Lys
               215                 220                 225
Glu His Pro Ile Phe Phe Met Phe Ile Gln Ile Ala Val Ile Ala
               230                 235                 240
Ile Phe Lys Ser Tyr Pro Thr Val Gly Asp Val Ala Leu Tyr Met
               245                 250                 255
Ala Phe Phe Pro Val Trp Asn His Leu Tyr Arg Phe Leu Arg Asn
               260                 265                 270
Ile Phe Val Leu Thr Cys Ile Ile Ile Val Cys Ser Leu Leu Phe
               275                 280                 285
Pro Val Leu Trp His Leu Trp Ile Tyr Ala Gly Ser Ala Asn Ser
               290                 295                 300
Asn Phe Phe Tyr Ala Ile Thr Leu Thr Phe Asn Val Gly Gln Ile
               305                 310                 315
Leu Leu Ile Ser Asp Tyr Phe Tyr Ala Phe Leu Arg Arg Glu Tyr
               320                 325                 330
Tyr Leu Thr His Gly Leu Tyr Leu Thr Ala Lys Asp Gly Thr Glu
               335                 340                 345
Ala Met Leu Val Leu Lys
               350
```

<210> 5
<211> 757
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte clone 1853196

```
<400> 5
Met Ser Glu Asn Ser Ser Asp Ser Asp Ser Ser Cys Gly Trp Thr
 1               5                   10                  15
Val Ile Ser His Glu Gly Ser Asp Ile Glu Met Leu Asn Ser Val
                20                  25                  30
Thr Pro Thr Asp Ser Cys Glu Pro Ala Pro Glu Cys Ser Ser Leu
                35                  40                  45
Glu Gln Glu Glu Leu Gln Ala Leu Gln Ile Glu Gln Gly Glu Ser
                50                  55                  60
Ser Gln Asn Gly Thr Val Leu Met Glu Glu Thr Ala Tyr Pro Ala
                65                  70                  75
Leu Glu Glu Thr Ser Ser Thr Ile Glu Ala Glu Glu Gln Lys Ile
                80                  85                  90
Pro Glu Asp Ser Ile Tyr Ile Gly Thr Ala Ser Asp Asp Ser Asp
                95                  100                 105
Ile Val Thr Leu Glu Pro Pro Lys Leu Glu Glu Ile Gly Asn Gln
                110                 115                 120
Glu Val Val Ile Val Glu Glu Ala Gln Ser Ser Glu Asp Phe Asn
                125                 130                 135
Met Gly Ser Ser Ser Ser Ser Gln Tyr Thr Phe Cys Gln Pro Glu
                140                 145                 150
Thr Val Phe Ser Ser Gln Pro Ser Asp Asp Glu Ser Ser Ser Asp
                155                 160                 165
Glu Thr Ser Asn Gln Pro Ser Pro Ala Phe Arg Arg Arg Arg Ala
                170                 175                 180
Arg Lys Lys Thr Val Ser Ala Ser Glu Ser Glu Asp Arg Leu Val
                185                 190                 195
Ala Glu Gln Glu Thr Glu Pro Ser Lys Glu Leu Ser Lys Arg Gln
                200                 205                 210
Phe Ser Ser Gly Leu Asn Lys Cys Val Ile Leu Ala Leu Val Ile
                215                 220                 225
Ala Ile Ser Met Gly Phe Gly His Phe Tyr Gly Thr Ile Gln Ile
                230                 235                 240
Gln Lys Arg Gln Gln Leu Val Arg Lys Ile His Glu Asp Glu Leu
                245                 250                 255
Asn Asp Met Lys Asp Tyr Leu Ser Gln Cys Gln Gln Glu Gln Glu
                260                 265                 270
Ser Phe Ile Asp Tyr Lys Ser Leu Lys Glu Asn Leu Ala Arg Cys
                275                 280                 285
Trp Thr Leu Thr Glu Ala Glu Lys Met Ser Phe Glu Thr Gln Lys
                290                 295                 300
Thr Asn Leu Ala Thr Glu Asn Gln Tyr Leu Arg Val Ser Leu Glu
                305                 310                 315
Lys Glu Glu Lys Ala Leu Ser Ser Leu Gln Glu Glu Leu Asn Lys
                320                 325                 330
Leu Arg Glu Gln Ile Arg Ile Leu Glu Asp Lys Gly Thr Ser Thr
                335                 340                 345
Glu Leu Val Lys Glu Asn Gln Lys Leu Lys Gln His Leu Glu Glu
                350                 355                 360
Glu Lys Gln Lys Lys His Ser Phe Leu Ser Gln Arg Glu Thr Leu
                365                 370                 375
Leu Thr Glu Ala Lys Met Leu Lys Arg Glu Leu Glu Arg Glu Arg
                380                 385                 390
Leu Val Thr Thr Ala Leu Arg Gly Glu Leu Gln Gln Leu Ser Gly
                395                 400                 405
Ser Gln Leu His Gly Lys Ser Asp Ser Pro Asn Val Tyr Thr Glu
                410                 415                 420
```

Lys Lys Glu Ile Ala Ile Leu Arg Glu Arg Leu Thr Glu Leu Glu
425 430 435
Arg Lys Leu Thr Phe Glu Gln Gln Arg Ser Asp Leu Trp Glu Arg
440 445 450
Leu Tyr Val Glu Ala Lys Asp Gln Asn Gly Lys Gln Gly Thr Asp
455 460 465
Gly Lys Lys Lys Gly Gly Arg Gly Ser His Arg Ala Lys Asn Lys
470 475 480
Ser Lys Glu Thr Phe Leu Gly Ser Val Lys Glu Thr Phe Asp Ala
485 490 495
Met Lys Asn Ser Thr Lys Glu Phe Val Arg His His Lys Glu Lys
500 505 510
Ile Lys Gln Ala Lys Glu Ala Val Lys Glu Asn Leu Lys Lys Phe
515 520 525
Ser Asp Ser Val Lys Ser Thr Phe Arg His Phe Lys Asp Thr Thr
530 535 540
Lys Asn Ile Phe Asp Glu Lys Gly Asn Lys Arg Phe Gly Ala Thr
545 550 555
Lys Glu Ala Ala Glu Lys Pro Arg Thr Val Phe Ser Asp Tyr Leu
560 565 570
His Pro Gln Tyr Lys Ala Pro Thr Glu Asn His His Asn Arg Gly
575 580 585
Pro Thr Met Gln Asn Asp Gly Arg Lys Glu Lys Pro Val His Phe
590 595 600
Lys Glu Phe Arg Lys Asn Thr Asn Ser Lys Lys Cys Ser Pro Gly
605 610 615
His Asp Cys Arg Glu Asn Ser His Ser Phe Arg Lys Ala Cys Ser
620 625 630
Gly Val Phe Asp Cys Ala Gln Gln Glu Ser Met Ser Leu Phe Asn
635 640 645
Thr Val Val Asn Pro Ile Arg Met Asp Glu Phe Arg Gln Ile Ile
650 655 660
Gln Arg Tyr Met Leu Lys Glu Leu Asp Thr Phe Cys His Trp Asn
665 670 675
Glu Leu Asp Gln Phe Ile Asn Lys Phe Phe Leu Asn Gly Val Phe
680 685 690
Ile His Asp Gln Lys Leu Phe Thr Asp Phe Val Asn Asp Val Lys
695 700 705
Asp Tyr Leu Arg Asn Met Lys Glu Tyr Glu Val Asp Asn Asp Gly
710 715 720
Val Phe Glu Lys Leu Asp Glu Tyr Ile Tyr Arg His Phe Phe Gly
725 730 735
His Thr Phe Ser Pro Pro Tyr Gly Pro Arg Ser Val Tyr Ile Lys
740 745 750
Pro Cys His Tyr Ser Ser Leu
755

<210> 6
<211> 1378
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte clone 037377

<400> 6
ggctcttcct tgcgtgtgcg tgcacgttgg gtgctggggg gtggagaccg gatcttcctc 60
gcttgggtac tttcctctcg gtgtgtgttt ctggccggag ccgtttcgcg acggcccggg 120
cgccccgccc caaccttcct tccctagacc ctcttctctc ccttcggctt ctctctttcg 180
gccggcgccg ccagttcctg ggcacaccc agaggtcccc ttctcgccgc cgcctgcaac 240
tgcgagggta gcccggggcc gcttggagtc gcccggacct gagaggctgc tgcactgggc 300
ctcagccagc cctccggatg ctggtgctgc catccccctg ccctcagcct ctggcatttt 360
cctccgttga gaccatggag ggccctcccc gtcggacttg ccgctcccca gaacctggac 420
cttcctcctc catcggatct ccccaggctt catctcctcc aaggcccaac cactacctgc 480
ttattgacac tcagggtgtc ccctacacag tgctggtgga cgaggagtca cagagggagc 540
caggggccag tggggctcca ggccagaaaa agtgctacag ctgccccgtg tgctcaaggg 600
tcttcgagta catgtcctac cttcagcgac acagcatcac ccactcggag gtaaagccct 660
tcgagtgtga catctgtggg aaggcattca agcgcgccag ccacttggca cggcaccatt 720
ccattcacct ggcgggtggt gggcggcccc acggctgccc gctctgccct cgccgcttcc 780
gggatgcggg tgagctggcc cagcacagcc gggtgcactc tggggaacgc ccgtttcagt 840
gtccacactg ccctcgccgc tttatggagc agaacacact gcagaaacac acgcggtgga 900
agcatccatg agccgggctg ccgggtgccc caggtaccac aggactttgc agggagcctg 960
gactcctgtc cagacacctg gtgagagcct gaggctggtg ttcagggccc tggacacaga 1020
cacagagcag ccgcatctca aggcagagc cctgcctgaa ggaggaatcc gtgagtaatc 1080
ttcaggtcct ccgtgttctg gagctgagat gggaatgagc ccctacacag aatggagtcc 1140
tctagcctaa agatatcagc tgttccatgg cagagccttg actggatgga ggtggggagt 1200
gtggtgtgta aagtctctgg cctcataaaa ggtggctgtg ggtcgtcagg aatctgcgcc 1260
atcttcctgg ggcttctgcg ctgttgttgg ggaagggacc ccagtcctgc cttccacccc 1320
ccaaccaggc ctgagactga tcaaacaata aacacgtttc ccactctgaa aaaaaaa 1378

<210> 7
<211> 1207
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte clone 162871

<400> 7
gggcggagtt gtgggagtgg aggaggaaga ggcggtaggg ggtacggggg ctggtcccag 60
aagatggcgg aggcggggga tttctggtag gtcctacttt aggacaagat gtggtaccgt 120
tgaagcgtca gtctttgatt cacagacagt tgagcttttc agctgggaag cctttccatt 180
tttttttttt ttaacggctt tctgaaccta tgaaaccatg gcagaaggag agacagagtc 240
acctgggccc aaaaagtgtg gcccatatat ttcatctgtc actagccaga gtgtgaactt 300
gatgattcga ggagtagtgc tattttttat tggagtattt cttgcattag tgttaaattt 360
acttcagatt cagagaaatg tgacgctctt tccacctgat gtgattgcaa gcatcttttc 420
ttctgcatgg tgggtacccc catgctgtgg cacggcttct gctgtgattg ggttattata 480
cccctgcatt gacagacatc taggagaacc acataaattt aaaagagagt ggtccagtgt 540
aatgcggtgt gtagcagtct ttgttggtat aaatcatgcc agtgctaaag tggatttcga 600
taacaacata cagttgtctc tcacactggc tgcactatcc attggactgt ggtggacttt 660
tgatagatct agaagtggtt ttggccttgg agtaggaatt gccttcttgg caactgtggt 720
cactcaactg ctagtatata atggtgttta ccaatataca tctccagatt tcctctatgt 780
tcgttcttgg ttaccatgta tattttttgc tggaggcata acaatgggaa acattggtcg 840
acaactggca atgtacgaat gtaaagttat cgcagaaaaa tctcatcagg aatgaagaag 900
gcaaaaaata tcttttgtac agaaaagcaa gatgaaaagg atgtgaaatg gtagatatac 960
caacaaaact tcagactgta aaattgccag gatgcagttt tccccttgat tggcgtgtgt 1020
gtatatatgg aataaatata tatatacaca cacacatatt actgcaatct gtgattgctt 1080
catctgtaaa tcagttgtaa acctttacat atttgactta aataactgta agatatatat 1140
gtactacatt aaaaagtgtt gattaataga tgaaattttt aaattaattt tttaaaacat 1200

gcccata 1207

<210> 8
<211> 1192
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte clone 236062

<400> 8
```
gtgagccgga gtcagaactg cgtctcgcga cccaggcgcg ggtttccgga ggacagccaa   60
caagcgatgc tgccgccgcc gtttcctgat tggttgtggg tggctacctc ttcgttctga  120
ttggccgcta gtgagcaaga tgctgagcaa gggtctgaag cggaaacggg aggaggagga  180
ggagaaggaa cctctggcag tcgactcctg gtggctagat cctggccacg cagcggtggc  240
acaggcaccc ccggccgtgg cctctagctc cctctttgac ctctcagtgc tcaagctcca  300
ccacagcctg cagcagagtg agccggacct gcggcacctg gtgctggtcg tgaacactct  360
gcggcgcatc caggcgtcca tggcacccgc ggctgccctg ccacctgtgc ctagcccacc  420
tgcagccccc agtgtggctg acaacttact ggcaagctcg gacgctgccc tttcagcctc  480
catggccagc ctcctggagg acctcagcca cattgagggc ctgagtcagg ctccccaacc  540
cttggcagac gaggggccac caggccgtag catcgggggga gcagcgccca gcctgggtgc  600
cttggacctg ctgggcccag ccactggctg tctactggac gatgggcttg agggcctgtt  660
tgaggatatt gacacctcta tgtatgacaa tgaactttgg gcaccagcct ctgagggcct  720
caaaccaggc cctgaggatg ggccgggcaa ggaggaagct ccggagctgg acgaggccga  780
attggactac ctcatggatg tgctggtggg cacacaggca ctggagcgac cgccggggcc  840
agggcgctga gccctcgtgc tggaatggtt gtctggtatc tgaactgagc ctgctggctg  900
gaccaactgt cctcgaaaag acacagctgg cttccctagt acagagaaca gggcttgggc  960
cactttggag agacagaatc tagtcctggg caacttcaca tccgtcctcc tgtctcaggg 1020
ctggcagggg gagcctggaa ttacccccta gtgatggaat gacagggtct ggtggggact 1080
gaattccctg gccctggggt catagcttgg gctgttcctt ctctgatacg ggaagagacc 1140
ccaatcagat ttttcaaatt aaagccagtc ctgggaaatc tcaaaaaaaa aa         1192
```

<210> 9
<211> 1631
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte clone 1596581

<400> 9
```
cggacgcgtg ggtgctggtg gtggctgtga cagtgcgggc ggccttgttc cgctccagtc   60
tggccgagtt catttccgag cgggtggagg tggtgtcccc actgagctct tggaagagag  120
tggttgaagg cctttcactg ttggacttgg gagtatctcc gtattctgga gcagtatttc  180
atgaaactcc attaataata tacctctttc atttcctaat tgactatgct gaattggtgt  240
ttatgataac tgatgcactc actgctattg ccctgtattt tgcaatccag gacttcaata  300
aagttgtgtt taaaaagcag aaactcctcc tagaactgga ccagtatgcc ccagatgtgg  360
ccgaactcat ccggacccct atggaaatgc gttacatccc tttgaaagtg gccctgttct  420
atctcttaaa tccttacacg attttgtctt gtgttgccaa gtctacctgt gccatcaaca  480
acaccctcat tgctttcttc attttgacta cgataaaagg cagtgctttc ctcagtgcta  540
tttttcttgc cttagcgaca taccagtctc tgtacccact caccttgttt gtcccaggac  600
tcctctatct cctccagcgg cagtacatac ctgtgaaaat gaagagcaaa gccttctgga  660
```

tcttttcttg ggagtatgcc atgatgtatg tgggaagcct agtggtaatc atttgcctct 720
ccttcttcct tctcagctct tgggatttca tccccgcagt ctatggcttt atactttctg 780
ttccagatct cactccaaac attggtcttt tctggtactt ctttgcagag atgtttgagc 840
acttcagcct cttctttgta tgtgtgtttc agatcaacgt cttcttctac accatcccct 900
tagccataaa gctaaaggag caccccatct tcttcatgtt tatccagatc gctgtcatcg 960
ccatctttaa gtcctacccg acagtggggg acgtggcgct ctacatggcc ttcttccccg 1020
tgtggaacca tctctacaga ttcctgagaa acatctttgt cctcacctgc atcatcatcg 1080
tctgttccct gctcttccct gtcctgtggc acctctggat ttatgcagga agtgccaact 1140
ctaatttctt ttatgccatc acactgacct tcaacgttgg gcagatcctg ctcatctctg 1200
attacttcta tgccttcctg cggcgggagt actacctcac acatggcctc tacttgaccg 1260
ccaaggatgg cacagaggcc atgctcgtgc tcaagtaggc ctggctggca cagggctgca 1320
tggacctcag gggctgtgg ggccagaagc tgggccaagc cctccagcca gagttgccag 1380
caggcgagtg cttgggcaga agaggttcga gtccagggtc acaagtctct ggtaccaaaa 1440
gggacccatg gctgactgac agcaaggcct atggggaaga actgggagct ccccaacttg 1500
gacccccacc ttgtggctct gcacaccaag gagcccccctc ccagacagga aggagaagag 1560
gcaggtgagc agggcttgtt agattgtggc tacttaataa atgtttttg ttatgaagtc 1620
taaaaaaaaa a 1631

<210> 10
<211> 3006
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte clone 1853196

<400> 10
ttggcgcccg agctgtgacc gccgccactg ggcagccag cacaatcggg cggaggtggc 60
gctgcccctt cagacctgaa agatgtctga aaattccagt gacagtgatt catcttgtgg 120
ttggactgtc atcagtcatg aggggtcaga tatagaaatg ttgaattctg tgacccccac 180
tgacagctgt gagcccgccc cagaatgttc atctttagag caagaggagc ttcaagcatt 240
gcagatagag caaggagaaa gcagccaaaa tggcacagtg cttatggaag aaactgctta 300
tccagctttg gaggaaacca gctcaacaat tgaggcagag gaacaaaaga tacccgaaga 360
cagtatctat attggaactg ccagtgatga ttctgatatt gttacccttg agccacctaa 420
gttagaagaa attggaaatc aagaagttgt cattgttgaa gaagcacaga gttcagaaga 480
ctttaacatg ggctcttcct ctagcagcca gtatactttc tgtcagccag aaactgtatt 540
ttcatctcag cctagtgatg atgaatcaag tagtgatgaa accagtaatc agcccagtcc 600
tgcctttaga cgacgccgtg ctaggaagaa gaccgtttct gcttcagaat ctgaagaccg 660
gctagttgct gaacaagaaa ctgaaccttc taaggagttg agtaaacgtc agttcagtag 720
tggtctcaat aagtgtgtta tacttgcttt ggtgattgca atcagcatgg gatttggcca 780
tttctatggc acaattcaga ttcagaagcg tcaacagtta gtcagaaaga tacatgaaga 840
tgaattgaat gatatgaagg attatctttc ccagtgtcaa caggaacaag aatcttttat 900
agattataag tcattgaaag aaaatcttgc aaggtgttgg acacttactg aagcagagaa 960
gatgtccttt gaaactcaga aaacgaacct tgctacagaa aatcagtatt taagagtatc 1020
cctggagaag gaagaaaaag ccttatcctc attacaggaa gagttaaaca aactaagaga 1080
acagattaga atattggaag ataaagggac aagtactgaa ttagttaaag aaaatcagaa 1140
acttaagcag catttggaag aggaaaagca gaaaaaacac agctttctta gtcaaaggga 1200
gactctgttg acagaagcaa agatgctaaa gagagaactg gagagagaac gactagtaac 1260
tacggcttta agggggggaac tccagcagtt aagtggtagt cagttacatg gcaagtcaga 1320
ttctcccaat gtatatactg aaaaaaagga aatagcaatc ttacgggaaa gactcactga 1380
gctggaacgg aagctaacct tcgaacagca gcgttctgat ttgtgggaaa gattgtatgt 1440
tgaggcaaaa gatcaaaatg gaaaacaagg aacagatgga aaaaagaaag gggcagaggg 1500
aagccacagg gctaaaaata agtcaaagga aacatttttg ggttcagtta aggaaacatt 1560
tgatgccatg aagaattcta ccaaggagtt tgtaaggcat cataaagaga aaattaagca 1620

ggctaaagaa gctgtgaagg aaaatctgaa aaaattctca gattcagtta aatccacttt 1680
cagacacttt aaagatacca ccaagaatat ctttgatgaa aagggtaata aaagatttgg 1740
tgctacaaaa gaagcagctg aaaaaccaag aacagttttt agtgactatt tacatccaca 1800
gtataaggca cctacagaaa accatcataa tagaggccct actatgcaaa atgatggaag 1860
gaaagaaaag ccagttcact ttaaagaatt cagaaaaaat acaaattcaa agaaatgcag 1920
tcctgggcat gattgtagag aaaattctca ttctttcaga aaggcttgtt ctggtgtatt 1980
tgattgtgct caacaagagt ccatgagcct ttttaacaca gtggtgaatc ctataaggat 2040
ggatgaattt agacagataa ttcaaaggta catgttaaaa gaactggata ctttttgtca 2100
ctggaacgaa cttgatcagt tcatcaataa gtttttccta aacggtgtct ttatacatga 2160
tcagaagctc ttcactgact ttgttaatga tgttaaagat tatcttagaa acatgaagga 2220
atatgaagta gataatgatg gagtatttga gaagttggat gaatatatat atagacactt 2280
ctttggtcac actttttccc ctccatatgg acccaggtcg gtttacataa aaccgtgtca 2340
ttacagtagt ttgtaacatt tgtagattgg atagcatttt tatgatttga tgagtttctt 2400
gtaaggttac cgtttctaag agttgtgctt tatggccact gagagaattc agaataaatt 2460
gaaagatgga gtctaaaaat tattagctgt tacaaatgga acatttcatt ataacgtgat 2520
cactttgact tgagcaaatg gtttaatttt tatcttaaaa atcagttaag aatatataaa 2580
atcctacttt ggccaagttt gtttcttttc attatagttt atatgaaaag atcaccttaa 2640
gtgaaattat tttcctttaa tcttttatgt atttattcac ttttggaagc taggaatgag 2700
caacacaaat tttactctga agtcagaaga gctcatatat aataattcta atgtcccacc 2760
tattttcact tgtccattcc atgtaccagc ttagttatga tacttagtca cataattatc 2820
tttgataaag gtagaggcac aaagaggcaa actaagcaag tcaaattcta atgtgtgtac 2880
ttcataataa ttttttatcc attttcatct ttatattctg taacatgaaa cttacctaat 2940
cttcaaatgt tagcttcatt ttttaccttt gaaatactta atctttctga ataaatataa 3000
tgtgta 3006

<210> 11
<211> 684
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 269, 285, 295, 312, 366, 375, 378, 397, 406, 428, 495, 501, 503
<221> unsure
<222> 586, 592, 610, 613, 643
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 108390F1

<400> 11
cagtatatac attgggagaa tctgacttgc catgtaactg actaccactt aactgctgga 60
gttccccccct taaagccgta gttactagtc gttctctctc cagttctctc tttagcatct 120
ttgcttctgt caacagagtc tccctttgac taagaaagct gtgtttttc tgcttttcct 180
cttccaaatg ctgcttaagt ttctgatttt ctttaactaa ttcagtactt gtccctttat 240
cttccaatat tctaatctgt tctcttagnt tgtttaactc ttccngtaat gaggntaagg 300
ctttttcttc cntctccagg gatactctta aatactgatt ttctgtagca aggttcgttt 360
ctgagnttca aaggncanct tctctgcttc agtaagngtc caacancttg caagatttct 420
ttcaatgnct tataatctat aaaagttctt gttcccgttg acacggggaa ggtaatcctc 480
atatcatcaa ttcancttca ngnatctttc tgactaactg ttgacggttc tgaatctgaa 540
tgtgccatag gaatggccaa atcccatgct gattgcaatc accaangcaa gnataacaca 600
cttattgggn ccnctactga actgacggtt actcaactcc ttnggagggt cagttcttgt 660
tcagcaacta gccggtcttc agat 684

<210> 12
<211> 416
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte clone 1211009R1

<400> 12
aagaattcta ccaaggagtt tgtaaggcat cataaagaga aaattaagca ggctaaagaa 60
gctgtgaagg aaaatctgaa aaaattctca gattcagtta aatccacttt cagacacttt 120
aaagatacca ccaagaatat ctttgatgaa aagggtaata aaagatttgg tgctacaaaa 180
gaagcagctg aaaaaccaag aacagttttt agtgactatt tacatccaca gtataaggca 240
cctacagaaa accatcataa tagaggccct actatgcaaa atgatggaag gaaagaaaag 300
ccagttcact ttaaagaatt cagaaaaaat acaaattcaa agaaatgcag tcctgggcat 360
gattgtagag aaaattctca ttctttcaga aaggcttgtt ctggtgtatt tgattg 416

<210> 13
<211> 609
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 25, 152, 166, 169, 173, 174, 180, 183, 186, 192, 193, 198, 200
<221> unsure
<222> 205, 220, 230, 233, 236, 243, 246, 251, 285, 307, 309, 310, 317
<221> unsure
<222> 319, 329, 344, 345, 377, 475, 485, 556, 573, 583, 594
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 1211009T1

<400> 13
aagaacatta tatttattca gaaanattaa gtatttcaaa ggtaaaaaat gaagctaaca 60
tttgaagatt aggtaagttt catgttacag aatataaaga tgaaaatgga taaaaaatta 120
ttatgaagta cacacattag aatttgactt gnttagtttg cctctntgng ccnntacctn 180
tancanaggt anntatgngn ctaantatca taactaagcn ggtacatggn atnganaagt 240
ganaanaggt nggacattag aaattattat atatgagctc ttctnacttc agagtaaaat 300
ttgtgtngnn cattccnanc ttccaaaant gaataaatac atannagatt aaaggaaaat 360
aatttcactt aaggtgntct tttcatataa actataatga gaagaaacaa acttggccaa 420
agtaggattt tatatattct taactgattt ttaagataga aaattaaacc atttnctcaa 480
gtcanagtga tcacgttata atgaaatgtt ccatttgtaa cagctaataa tttttagact 540
ccatctttca atttantctg aattctctca gtngccataa agncaactct tagnaacggt 600
accttcaag 609

<210> 14
<211> 189
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte clone 1352052H1

<400> 14
cttcacatcc gtcctcctgt ctcagggctg gcagggggag cctggaatta ccccctagtg 60
atggaatgac agggtctggt ggggactgaa ttccctggcc ctggggtcat agcttgggct 120
gttccttctc tgatacggga agagacccca atcagatttt tcaaattaaa gccagtcctg 180
ggaaatctc 189

<210> 15
<211> 473
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 34, 59, 60, 134, 168, 311, 314, 344, 347, 354, 364, 391, 393, 401
<221> unsure
<222> 407, 413, 416, 426, 445, 446, 447, 453, 454, 459, 471
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 1391767F1

<400> 15
gaaaaaaagg aaatagcaat cttacgggaa agantcactg agctggaacg gaagctaann 60
ttcgaacagc agcgttctga tttgtgggaa agattgtatg ttgaggcaaa agatcaaaat 120
ggaaaacaag gaanagatgg aaaaaagaaa gggggcagag gaagccanag ggctaaaaat 180
aagtcaaagg aaacattttt gggttcagtt aaggaaacat ttgatgccat gaagaattct 240
accaaggagt ttgtaaggca tcataaagag aaaattaagc aggctaaaga agctgtgaag 300
gaaaatctga naanattctc agattcagtt aaatccactt tccnggnact ttanagtacc 360
cccnagggta tctttgatga aaagggtaat nanagtttgg ngctacnaaa gangcnagct 420
gaaaanccag gacagttttt agggnnntat tgnnatccnc agtataaggc ncc 473

<210> 16
<211> 529
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 119, 501
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 1477338F1

<400> 16
ccccagatgt ggccgaactc atccggaccc ctatggaaat gcgttacatc cctttgaaag 60
tggccctgtt ctatctctta aatccttaca cgattttgtc ttgtgttgcc aagtctacnt 120
gtgccatcaa caacaccctc attgctttct tcattttgac tacgataaaa ggcagtgctt 180
tcctcagtgc tatttttctt gccttagcga cataccagtc tctgtaccca ctcaccttgt 240
ttgtcccagg actcctctat ctcctccagc ggcagtacat acctgtgaaa atgaagagca 300
aagccttctg gatcttttct tgggagtatg ccatgatgta tgtgggaagc ctagtggtaa 360
tcatttgcct ctccttcttc cttctcagct cttgggattt catccccgca gtctatggct 420
ttatactttc tgttccagat ctcactccaa acattggtct tttctggtac ttctttgcag 480
agatgtttga gcacttcagc ntcttctttg tatgtgtgtt cagatcaac 529

<210> 17
<211> 581
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 372, 374, 445
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 1520634F1

<400> 17
gccatccccc tgccctcagc ctctggcatt ttcctccgtt gagaccatgg agggccctcc 60
ccgtcggact tgccgctccc cagaacctgg accttcctcc tccatcggat ctccccaggc 120
ttcatctcct ccaaggccca accactacct gcttattgac actcagggtg tcccctacac 180
agtgctggtg gacgaggagt cacagaggga gccaggggcc agtggggctc caggccagaa 240
aaagtgctac agctgccccg tgtgctcaag ggtcttcgag tacatgtcct accttcagcg 300
acacagcatc acccactcgg aggtaaagcc cttcgagtgt gacatctgtg ggaaggcatt 360
caagcgcgcc ancnacttgg cacggcacca ttccattcac ctggcgggtg gtgggcggcc 420
ccacggctgc ccgctctgcc ctcgncgttc cgggatgcgg gtgagctggc ccagcacagc 480
cgggtgcact ctggggaacg cccgtttcag tgtcacactg cctcgccgtt tatggagaga 540
acacactgca gaaacacacg ggtggaagca tccatgagcg g 581

<210> 18
<211> 637
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 462, 485, 510, 514, 550, 562, 602, 617, 622, 625, 629, 636
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 1525569F6

<400> 18
cagtaatcag cccagtcctg cctttagacg acgccgtgct aggaagaaga ccgtttctgc 60

ttcagaatct gaagaccggc tagttgctga acaagaaact gaaccttcta aggagttgag 120
taaacgtcag ttcagtagtg gtctcaataa gtgtgttata cttgctttgg tgattgcaat 180
cagcatggga tttggccatt tctatggcac aattcagatt cagaagcgtc aacagttagt 240
cagaaagata catgaagatg aattgaatga tatgaaggat tatctttccc agtgtcaaca 300
ggaacaagaa tcttttatag attataagtc attgaaagaa aatcttgcaa ggtgttggac 360
acttactgaa gcagagaaga tgtcctttga aactcagaaa acgaaccttg ctaccagaaa 420
atcagtattt aagagtatcc ttggagaagg aagaaaaagc cntatcctca ttaccaggga 480
agagntaaac aaacttaaga ggaccagttn gganattgga agataaaggg gacaagtact 540
gaattagttn aaggaaaatc cngaaacttt aagcagcctt tggaagaggg aaagccggaa 600
anacaccagc tttcctnagt cnaangggng accctnt 637

<210> 19
<211> 187
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 13, 19, 21
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 1554775H1

<400> 19
gggcggagtt gtnggagtng nggaggaaga ggcggtaggg ggtacggggg ctggtcccag 60
aagatggcgg aggcggggga tttctggtag gtcctacttt aggacaagat gtggtaccgt 120
tgaagcgtca gtctttgatt cacagacagt tgagcttttc agctgggaag cctttccatt 180
ttttttt 187

<210> 20
<211> 499
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 406, 435
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 1596581F6

<400> 20
aaaaagcaga aactcctcct agaactggac cagtatgccc cagatgtggc cgaactcatc 60
cggaccccta tggaaatgcg ttacatccct ttgaaagtgg ccctgttcta tctcttaaat 120
ccttacacga ttttgtcttg tgttgccaag tctacctgtg ccatcaacaa caccctcatt 180
gctttcttca ttttgactac gataaaaggc agtgctttcc tcagtgctat tttcttgcc 240
ttagcgacat accagtctct gtacccactc accttgtttg tcccaggact cctctatctc 300

ctccagcggc agtacatacc tgtgaaaatg aagagcaaag ccttctggat cttttcttgg 360
gagtatgcca tgatgtatgt gggaagccta gtggtaatca tttgcntctc cttcttcctt 420
ctcagctctt ggganttcat ccccgcagtc taatggctta tactttctgt tccagatctc 480
atccaaacat tgggtcttt 499

<210> 21
<211> 287
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 122, 144, 266, 273
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 1596581T1

<400> 21
ggcttggccc agcttctggc cccacagccc cctgaggtcc atgcagccct gtgccagcca 60
ggcctacttg agcacgagca tggcctctgt gccatccttg gcggtcaagt agaggccatg 120
tntgaggtag tactcccgcc gcangaaggc atagaagtaa tcagagatga gcaggatctg 180
cccaacgttg aaggtcagtg tgatggcata aaagaaatta gagttggcac ttcctgcata 240
aatccagagg tgccacagga cagggnagaa cangggacag acgattt 287

<210> 22
<211> 579
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 22, 25, 32, 106, 123, 126, 135, 208, 216, 219, 234, 236, 263, 271
<221> unsure
<222> 282, 287, 292, 358, 360, 363, 365, 379, 412, 441, 452, 459, 483
<221> unsure
<222> 485, 499, 500
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 162871X4

<400> 22
ctaaagaaga gcggtagggg gnccnggggc tngtcccaga aagtatggcg gaggcggggg 60
atttctggta ggtcctactt taggacaaga tgtggtaccg ttgaancgtc agtctttgat 120
tcncanacag ttganctttt cagctgggaa gcctttccat tttttttttt aacggctttc 180
tgaacctatg aaaccatggc aaaagganaa acaaantcnc ctgggcccaa aaantntggc 240
ccatatattt catctgtcac tanccaaatt ntgaacttga tnattcnagg antattgcta 300
ttttttattg gagtatttct tgcattagtg ttaaatttac ttcaaattca aaaaaatntn 360
acncnctttc cacctgatnt gattgcaagc atcttttctt ctgcatgctg tnattgggtt 420
attataccc tgcattaaca nacatctagg anaaccacnt aaatttaaaa aaaagtggtc 480

cantntaatg cggtgtgtnn cagtctttgt tggtataaat catgccagtg ctaaagtgga 540
tttcgataac aacatacagt tgtctctcac actggcgca 579

<210> 23
<211> 250
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 8, 17, 24, 27, 33, 36, 43, 246
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 162871X92

<400> 23
ggaaaggnca agggaanggg gtanggncgg tgntcnaaga aantggcgga ggcgggggat 60
ttctgctgtg attgggttat tataccccctg cattgacaga catctaggag aaccacataa 120
atttaaaaga gagcggtcca gtgtaatgcg gtgtgtagca gtctttgttg gtataaatca 180
tgccagtgct aaagtggatt tcgataacaa catacagttg tctctcacac tggctgcact 240
atcttnaaaa 250

<210> 24
<211> 250
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 8
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 1658067H1

<400> 24
cgacagtngg ggacgtggcg ctctacatgg ccttcttccc cgtgtggaac catctctaca 60
gattcctgag aaacatcttt gtcctcacct gcatcatcat cgtctgttcc ctgctcttcc 120
ctgtcctgtg gcacctctgg atttatgcag gaagtgccaa ctctaatttc ttttatgcca 180
tcacactgac cttcaacgtt gggcagatcc tgctcatctc tgattacttc tatgccttcc 240
tgcggcggga 250

<210> 25
<211> 736
<212> DNA
<213> Homo sapiens

<220>

<221> unsure
<222> 419, 435, 453, 462, 463, 471, 476, 513, 516, 563, 585, 586, 597
<221> unsure
<222> 611, 618, 652, 661, 680, 684, 685, 692, 693, 701, 714, 725, 731
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 1706512F6

<400> 25
```
atcagaagct cttcactgac tttgttaatg atgttaaaga ttatcttaga aacatgaagg  60
aatatgaagt agataatgat ggagtatttg agaagttgga tgaatatata tatagacact 120
tctttggtca cactttttcc cctccatatg gacccaggtc ggtttacata aaaccgtgtc 180
attacagtag tttgtaacat ttgtagattg gatagcattt ttatgatttg atgagtttct 240
tgtaaggtta ccgtttctaa gagttgtgct ttatgggcac tgagagaatt ccagaataaa 300
ttgaaagatg ggagtcctaa aaatttaatt agccggttac caaatgggga cctttttccat 360
tagtaacggt gattccacct ttggaccttt gaggccaaat gggtttaaat ttttttaanc 420
ccttaaaaaa atccnggttt aaaggaatta ttnttaaaga anncccccacc nttttngggc 480
ccaaggtttt ggttttccct ttttccattt aanaanggtt ttaataatgg aaaaaaggat 540
tccacccttt aaaggtggga aantttaatt ttttccccct taaanncccct ttttaanggg 600
aatttaaatt ncccccttnct gggaagccca agggaatgga ggcccacccc cnaattttta 660
ncccccggaag gtccggaagn ggcnnectat annaataatt nccaaaggtc cccncccaat 720
tttcncctgg ncccat                                                 736
```

<210> 26
<211> 611
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 213, 223, 369, 406, 423, 469, 475, 490, 494, 498, 524, 548, 570
<221> unsure
<222> 574, 582, 584, 594, 597, 605, 607
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 1722946F6

<400> 26
```
attggcgccc gagctgtgac cgccgccact ggggcagcca gcacaatcgg gcggaggtgg  60
cgctgcccct tcagacctga aagatgtctg aaaattccag tgacagtgat tcatcttgtg 120
gttggactgt catcagtcat gaggggtcag atatagaaat gttgaattct gtgacccccca 180
ctgacagctg tgagcccgcc ccagaatgtt canctttaga gcnagaggag cttcaagcat 240
tgcagataga gcaaggagaa tgcagccaaa atggcacagt gcttatggaa gaaactgctt 300
atccagcttt ggaggaaacc agctcaacaa ttgaggcaga ggaacaaaag atacccgaag 360
acagtatcna tattggaact gccagtggtg attctgatat tgttancccct tgagccacta 420
agnttagaag gaattgggga tccaagaagt tgtcattgtt gaagaaagnc caagntccgg 480
agacttttan catngggntc ttcctctagc agccagtata cttntctgtt cagcccagaa 540
aactggantt tcatcttcag cctaatgacn gtgnaatcaa gntngtgatg gaanccngtt 600
attcngnccc c                                                      611
```

<210> 27
<211> 592
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 94, 104, 149, 167, 215, 226, 232, 275, 298, 301, 312, 333, 362
<221> unsure
<222> 364, 367, 376, 391, 392, 395, 412, 415, 419, 429, 435, 443, 449
<221> unsure
<222> 452, 462, 463, 464, 466, 467, 468, 470, 476, 485, 489, 492, 502
<221> unsure
<222> 514, 529, 533, 541, 550, 557, 558, 567, 572, 574, 577, 580
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 1853196F6

<400> 27
```
ctttcagaaa ggcttgttct ggtgtatttg attgtgctca acaagagtcc atgagccttt  60
ttaacacagt ggtgaatcct ataaggatgg atgnatttag acanataatt caaaggtaca 120
tgttaaaaga actggatact ttttgtcant ggaacgaact tgatcanttc atcaataagt 180
ttttcctaaa cggtgtcttt atacatgatc agaanctctt cactgncttt gntaatgatg 240
ttaaagatta tcttagaaac atgaaggata tgaantagat aatgatggag tatttgcnaa 300
nttggatgga tntatatata gacacttctt tgntcacact ttttcccctc catatgggcc 360
cngntcngtt tacatnaaac cgtgtcttac nntantttgt aacatttgta gntgnatanc 420
atttttaant ttgangagtt tcntgtaang tnacggttcc annngnnntn ctttanagcc 480
accanagana antcggataa antgaaagta gggntccaaa attattaant gtnccaatag 540
nactttcctn ataaagnngt caccttngct tnanccnatn ggtttaattt tt         592
```

<210> 28
<211> 447
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte clone 2238411F6

<400> 28
```
tttgggcacc agcctctgag ggcctcaaac caggccctga ggatgggccg ggcaaggagg  60
aagctccgga gctggacgag gccgaattgg actacctcat ggatgtgctg gtgggcacac 120
aggcactgga gcgaccgccg gggccagggc gctgagccct cgtgctggaa tggttgtctg 180
gtatctgaac tgagcctgct ggctggacca actgtcctcg aaaagacaca gctggcttcc 240
ctagtacaga gaacagggct tgggccactt tggagagaca gaatctagtc ctgggcaact 300
tcacatccgt cctcctgtct cagggctggc agggggagcc tggaattacc ccctagtgat 360
ggaatgacag ggtctggtgg ggactgaatt ccctggccct ggggtcatag cttgggctgt 420
tccttctctg atacgggaag agacccc                                    447
```

<210> 29
<211> 247

<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 234
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 2312928H1

<400> 29
```
tgctggtggt ggctgtgaca gtgcgggcgg ccttgttccg ctccagtctg gccgagttca  60
tttccgagcg ggtggaggtg gtgtccccac tgagctcttg gaagagagtg gttgaaggcc 120
tttcactgtt ggacttggga gtatctccgt attctggagc agtatttcat gaaactccat 180
taataatata cctctttcat ttcctaattg actatgctga attggtgttt atgntaactg 240
atgcact                                                           247
```

<210> 30
<211> 190
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 162, 163
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 3015795H1

<400> 30
```
acttcacatc cgtcctcctg tctcagggct ggcaggggga gcctggaatt accccctagt  60
gatggaatga cagggtctgg tggggactga attccctggc cctggggtca tagcttgggc 120
tgttccttct ctgatacggg aagagacccc aatcagattt tnnaaattaa agccagtcct 180
gggaaatctc                                                        190
```

<210> 31
<211> 253
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 121
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature

<223> Incyte clone 3231214H1

<400> 31
gtttcagatc aacgtcttct tctacaccat cccccttagcc ataaagctaa aggagcaccc 60
catcttcttc atgtttatcc agatcgctgt catcgccatc tttaagtcct acccgacagt 120
nggggacgtg gcgctctaca tggccttctt ccccgtgtgg aaccatctct acagattcct 180
gagaaacatc tttgtcctca cctgcatcat catcgtctgt tccctggctc ttccctgtcc 240
tgtggcacct ctg 253

<210> 32
<211> 273
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 88
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 3985439H1

<400> 32
gtcttccttg cgtgtgcgtg cacgttgggt gctgggggt ggagaccgga tctatcctcg 60
cttgggtact ttcctctcgg tgtgtgtntc tggccggagc cgtttcgcga cggcccgggc 120
gccccgcccc aaccttcctt ccctagaccc tcttctctcc cttcggcttc tctctttcgg 180
ccggcgccgc cagttcctgg ggcacaccca gaggtcccct tctcgccgcc gcctgcaact 240
gcgagggtag cccggggccg cttggagtcg ccc 273

<210> 33
<211> 618
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 190, 336, 351, 413, 420, 423, 432, 441, 449, 454, 462, 510, 520
<221> unsure
<222> 524, 530, 552, 555, 557, 560, 561, 569, 574, 584, 585, 594, 596
<221> unsure
<222> 611, 614
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 403002R6

<400> 33
tgctggtgga cgaggagtca cagagggagc caggggccag tggggctcca ggccagaaaa 60
agtgctacag ctgccccgtg tgctcaaggg tcttcgagta catgtcctac cttcagcgac 120

acagcatcac ccactcggag gtaaagccct tcgagtgtga catctgtggg aaggcattca 180
agcgcgccan ccacttggca cggcaccatt ccattcacct ggcgggtggt gggcggcccc 240
acggctgccc gctctgccct cgccgcttcc gggatgcggg tgagtggccc aagcacagcc 300
gggtgcactc tggggaacgc ccgtttcagt gtcaanactg ccttcgccgg ntttaatgga 360
gcagaacaca attgcagaaa acaacaccgc ggttggaaag catcccattg aanccggggn 420
ttnccgggtt tnccccaagg ntaccaaang gaantttttc anaggggaac ccttgaaatt 480
ccctgttcca aaaaaacctt ggttaaaaan ccctaaaggn tggnttttn aggggccttg 540
gaaaaacagg ancananggn nagcgggant tttnaaaggg aaanncccctt gccnanaagg 600
gggaatcccg naantaat 618

<210> 34
<211> 297
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte clone 510407R6

<400> 34
tgagtaatct tcaggtcctc cgtgttctgg agctgagatg ggaatgagcc cctacacaga 60
atggagtcct ctagcctaaa gatatcagct gttccatggc agagccttga ctggatggag 120
gtggggagtg tggtgtgtaa agtctctggc ctcataaaag gtggctgtgg gtcgtcagga 180
atctgcgcca tcttcctggg gcttctgcgc tgttgttggg gaagggaccc cagtcctgcc 240
ttccaccccc caaccaggcc tgagactgat caaacaataa acacgtttcc cactctg 297

<210> 35
<211> 239
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 91
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte clone 3590729H1

<400> 35
ggcgagtgct tgggcagaag aggttcgagt ccagggtcac aagtctctgg taccaaaagg 60
gacccatggc tgactgacag caaggcctat ngggaagaac tgggagctcc ccaacttgga 120
cccccacctt gtggctctgc acaccaagga gccccctccc agacaggaag gagaagaggc 180
aggtgagcag ggcttgttag attgtggcta cttaataaat gttttttgtt atgaagtct 239